PhishGuru: A System for Educating Users about Semantic Attacks

Ponnurangam Kumaraguru

CMU-ISR-09-106
April 14, 2009

School of Computer Science
Institute for Software Research
Carnegie Mellon University
Pittsburgh PA 15213

Thesis Committee
Lorrie Cranor (Chair)
Jason Hong
Vincent Aleven
Rahul Tongia
Alessandro Acquisti

Submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

©2009 Ponnurangam Kumaraguru
All rights reserved

This research was partially funded by National Science Foundation grant number CCF-0524189 entitled "Supporting Trust Decisions," Army Research Office grant number DAAD19-02-1-0389, Fundaçã paraa Ciěnciae Tecnologia (FCT) Portugal under a grant from the Information and Communications Technology Institute (ICTI) at CMU, and Carnegie Mellon University CyLab. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect those of the funding agencies.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **14 APR 2009**	2. REPORT TYPE	3. DATES COVERED **00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE **PhishGuru: A System for Educating Users about Semantic Attacks**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Carnegie Mellon University,School of Computer Science,Institute for Software Research,Pittsburgh,PA,15213**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**198**	

Keywords: Social engineering, semantic attack, phishing, learning science, human computer interaction, trust, design and implementation, and real-world studies

Abstract

Online security attacks are a growing concern among Internet users. Currently, the Internet community is facing three types of security attacks: physical, syntactic, and semantic. Semantic attacks take advantage of the way humans interact with computers or interpret messages. There are three major approaches to countering semantic attacks: silently eliminating the attacks, warning users about the attacks, and training users not to fall for the attacks. The existing methods for silently eliminating the attack and warning users about the attack are unlikely to perform flawlessly; furthermore, users are the weakest link in these attacks, it is essential that user training complement other methods. Most existing online training methodologies are less successful because: (1) organizations that create and host training materials expect users to proactively seek out such material themselves; (2) these organizations expect users to have some knowledge about semantic attacks; and (3) the training materials have not been designed with learning science principles in mind.

The goal of this thesis is to show that computer users trained with an embedded training system – one grounded in the principles of learning science – are able to make more accurate online trust decisions than users who read traditional security training materials, which are distributed via email or posted online. To achieve this goal, we focus on "phishing," a type of semantic attack. We have developed a system called "PhishGuru" based on embedded training methodology and learning science principles. Embedded training is a methodology in which training materials are integrated into the primary tasks users perform in their day–to–day lives. In contrast to existing training methodologies, the PhishGuru shows training materials to users through emails at the moment ("teachable moment") users actually fall for phishing attacks.

We evaluated the embedded training methodology through laboratory and field studies. Real-world experiments showed that people trained with PhishGuru retain knowledge even after 28 days. PhishGuru training does not decrease users' willingness to click on links in legitimate messages. PhishGuru is also being used in a real-world implementation of the Anti-Phishing Working Group Landing Page initiative. The design principles established in this thesis will help researchers develop systems that can train users in other risky online situations.

Dream, Dream, Dream! Dreams transform into thoughts and thoughts into actions.

∼ Dr. A. P. J. Abdul Kalam, Former President of India

Acknowledgments

My advisor, Lorrie Faith Cranor, has brought about a lot of changes in me as a researcher, which will benefit me throughout my future. She has always been supportive of my research and activities at Carnegie Mellon University (CMU). All of the research I have conducted in completing my thesis wouldn't have been possible without her valuable input and constant encouragement. I am extremely fortunate to have her as my advisor. My thesis would not have been possible without the support, liberty, and motivation of my advisor for the last 5 years.

I had continuous support in hashing out ideas for my thesis from my committee members: Jason Hong, Vincent Aleven, Rahul Tongia, and Alessandro Acquisti. I thank Jason for his insightful comments on my papers, talks, and other documents. His advise on how to plan a research career has been very valuable and his principle "Keep it Simple" has influenced me a lot. I thank Vincent for his advice to "be skeptical" (in a good sense) about research results. He also always motivated me to be very strong on research methods. Rahul has influenced me in many ways, specifically, on how to conduct research that has a "social impact." Alessandro has provided me very good insights during my initial years as a Ph.D. student, particularly while I was pinning down my research problem, and I have learned a lot from his "critical" comments on writing research results. All my committee members have offered innumerable comments, suggestions, and feedback about the research presented in my thesis.

I would like to thank Raj Reddy for his vision and the trust he showed in my research capabilities. I would also like to thank Rajeev Sangal and Vasudeva Varma of the International Institute of Information Technology, Hyderabad, and Sandip Deb of SlashSupport, and Nandkumar Saravade and the late Sunil Mehta of the National Association of Software and Services Companies (NASSCOM) for motivating me to conduct applied research.

I also express my sincere gratitude to all faculty, students and members of the Computation, Organizations, and Society Program. I was very fortunate to be associated with a very active research group, Supporting Trust Decision; I would like to thank the members of this research group and the members of CyLab Usable Privacy and Security (CUPS) Lab. In particular, I got a lot out of discussions with my colleagues in CUPS lab: Serge Egelman, Patrick Kelley, Robert McGuire, Aleecia McDonald, Rob Reeder, Steve Sheng, Janice Tsai, and Kami Vaniea. They have made my research life at CMU great fun.

I would like to thank Mary Ann Blair, Theodore Pham, Wiam Younes, and others at the Information Security Office for their support. I would also like to thank Laura Mather, Nicole Loffredo, David Shroyer, and others at Anti-Phishing Working Group for their support and feedback.

Some of the system administrators and administrative staff members at the Institute for Software Research and CyLab have made my stay at CMU very pleasant: Emanuel Bowes, Chris Dalansky, Monika DeReno, Connie Herold, Helen Higgins, Karen Lindenfelser, Sherice Livingston, Jennifer Lucas, Janine Pischke, Ed Walter, and Linda Whipkey. I would like to extend my thanks to them.

I would also like to thank Linda Gentile and Genevieve Cook of the Office of International Education for taking care of my visa and travel issues promptly. I would also like to thank Julian Cantella, Edward Barr, Sarah Jameson, and Meg Voorhis for their feedback on the presentation and language of the various documents I have written during my stay at CMU.

This thesis would not have been possible without the support of my parents and relatives, my wife

Pranjali, and my son Aarav. Pranjali has played the role of friend, philosopher, and guide; without her, none of the work mentioned in my thesis would have come true.

Without friends it is really difficult for any one to achieve anything; it has not been any different with me. I had many wonderful friends who have made my graduate life a memorable one: Vamshi Ambati, Surendra Bagde, Michael Benisch, Aditya Bhave, Ajinkya Bhave, Chirantan Chatterjee, George Davis, Serge Egelman, Hemant Gogineni, Rohit Kumar, Anuj Kumar, Kishore Prahallad, Lavanya Prahallad, and Naveena Yanamala. Special thanks to Justin Cranshaw, Sharique Hasan and Steve Sheng for bearing with me on different discussions that I have had with them which have culminated in parts of my thesis.

I had a great opportunity to work with some brilliant undergraduate and masters students: Hajin Choi, Jerry Feng, Elizabeth Nunge, Yong Rhee, and Shelley Zhen.

Even though others have contributed to my thesis work, I am responsible for the content, conclusions, errors, or omissions in my thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Trust is an important component of human interactions. Individuals need some level of trust in other parties in order to interact with them in their personal and professional lives. Similarly, companies and organizations need to build trust among partners and customers in order to thrive in the marketplace. Online decisions are often trust-sensitive because many web-based transactions carry an element of uncertainty and risk [49, 127]. At the core of the uncertainty associated with Internet transactions is the problem of *incomplete* and, specifically, *asymmetric* information [3,190]: the reader of an email usually knows less than its sender about the accuracy of the information within the email or the sender's true identity; likewise, a visitor to a website knows less than the entity managing the site about how the personal data the visitor is asked to provide will be used. In an online scenario, the clues and signals we use in the offline world to fill in the gaps of incomplete information may be unavailable or deceptive.

As people conduct an increasing number of transactions using the Internet, Internet security concerns have also increased. Currently, the Internet community faces three types of security attacks: *physical*, where computers and electronics are targeted (e.g. power and data outages); *syntactic*, where operating logic and networks are targeted (e.g. security vulnerabilities in the Microsoft Windows operating system); and *semantic*, where the way users assign meaning to the content is targeted. A semantic attack is an attack that exploits human vulnerabilities. Rather than taking advantage of system vulnerabilities, semantic attacks take advantage of the way humans interact with computers or interpret messages [175, 176].

Since 2003, there has been a dramatic increase in a form of semantic attack known as phishing, in which victims get conned by spoofed emails or fraudulent websites. Phishing attacks exploit users' inability to distinguish legitimate company websites from fake ones. More specifically, phishers exploit the difference between what the system thinks the user is doing (*system model*) and what the users think the system is doing (*user mental model*) [138]. Emails are currently an important threat vector for exploiting this difference [93]. Phishers send out spoofed emails that look as if they were sent by trusted companies. These emails lead to spoofed websites that are similar or virtually identical to legitimate websites, luring people into disclosing sensitive information. Phishers use this information for criminal purposes such as identity theft, financial fraud, and corporate espionage [96, 116].

The number of phishing emails sent to users has increased notably over the last few years. Approximately 240 to 500 million phishing emails are now sent over the Internet each day [135, 180]. The number of phishing incidents reported to the Anti-Phishing Working Group (APWG) increased from 21 in Nov 2003 to 28,151 in June 2008. Seventy-three million U.S. adults said that they "definitely" or "think they" received an average of more than 50 phishing e-mails in 2005 [74]. The number of unique websites which have been phished is also increasing consistently [15].

The actual cost of phishing is difficult to calculate. Broadly, the cost of phishing can be classified as: *direct cost*, the cost directly incurred due to the phishing attack; *indirect cost*, the cost of handling customer support calls due to the phishing attack for an organization as well as the cost of the emotional stress for consumers; and *opportunity cost*, the cost incurred when distrust leads users to avoid using the Internet to do business and other financial transactions. In a survey of 5000 US consumers, Gartner found that nearly 30% of consumers changed their online banking behavior because of online attacks like phishing [74]. Gartner found that in 2007, 3.6 million adults lost $3.2 billion as a result of phishing attacks [160].

Developing countermeasures for phishing is a challenging problem because victims help attackers by giving away their credentials. It is also difficult to detect phishing websites and emails because they often look legitimate. In addition, users frequently ignore warning messages about phishing from anti-phishing tools [58, 67, 205].

A variety of strategies have been proposed to protect people from phishing. These strategies fall into three major categories: *silently eliminating the threat* by finding and taking down phishing websites, and by automatically detecting and deleting phishing emails; *warning users about the threat* through toolbars, browser extensions, and other mechanisms; and *training users not to fall for attacks*. There is no single silver bullet solution for the problem of phishing. We believe that these approaches are complementary. Specifically, automated detection systems should be used as the first line of defense against phishing attacks; however, since these systems are unlikely to perform flawlessly, they should be complemented by better user interfaces and user education programs that help people better recognize fraudulent emails and websites.

Most anti-phishing research has focused on solving the problem by eliminating the threat or warning users. However, little work has been done on educating people about phishing and other semantic attacks. Educating users about security is challenging, particularly in the context of phishing, because: (1) users are not motivated to read about security in general and therefore do not take time to educate themselves about phishing; (2) for most users, security is a secondary task (e.g.

one does not go to an online banking website to check the SSL implementation of the website, but rather to perform a banking transaction); and (3) it is difficult to teach people to make the right online trust decision without also increasing their concern level or tendency to misjudge non-threats as threats.

1.1 Thesis statement

Keeping the above challenges in mind, we address the problem of phishing user-education in this thesis. The thesis statement is:

> **Computer users trained using an embedded training system grounded in learning science theory are able to make more accurate online trust decisions than those who read traditional security training materials distributed via email or posted on web sites.**

1.2 Thesis contribution

This thesis is both timely and needed to reduce the negative consequences of semantic attacks on society. Results from this research can potentially help reduce the increasing number of people who fall for phishing and other semantic attacks. This research fits in the area of Computation, Organizations, and Society because phishing is a societal problem that can be solved by the collective efforts of computer science researchers, education researchers, lawyers, and organizations. This thesis work builds on existing knowledge in the fields of computer security and privacy, human computer interaction, learning science, and economics by building a system to help users make better online trust decisions. The design principles established in this thesis will help researchers develop systems that can train users in other risky online situations.

1.2.1 Real world impact

1. An implementation of PhishGuru (Anti-Phishing Working Group landing page), a system that we designed, is viewed world-wide approximately 500 times a day.

2. Anti-Phishing Phil, a game that we designed and evaluated, has been played over 100,000 times world-wide.

1.2.2 Theoretical

1. This research showed that computer users can be trained to make better online trust decisions if training materials are presented during their regular use of emails (PhishGuru).

2. This research showed that computer users can be trained to make better online trust decisions if the training materials are presented in a fun and interactive manner (Anti-Phishing Phil).

3. This research showed that computer users can be trained to make accurate online trust decisions if the training materials are grounded in learning science principles.

1.2.3 Design and development

1. We designed and developed a novel embedded training methodology to deliver phishing training materials to computer users.

2. We designed and developed a game that teaches computer users how to identify phishing URLs.

3. We implemented a variety of training interventions that use instructional design principles to address different phishing attack situations.

1.2.4 Experimental setup

1. We designed and conducted large phishing experiments (laboratory and real-world) to quantify the benefits of PhishGuru. Experimental design for evaluating phishing interventions is not trivial. Challenges include: (1) ensuring the privacy of participants; (2) delivering emails; (3) tracking users; (4) avoiding subject contamination; (5) creating the right incentive for participants to behave realistically; (6) running the study for a long time to measure retention; and (7) working with system administrators.

2. We designed and conducted experiments (laboratory and real-world) to quantify the benefits of Anti-Phishing Phil.

1.3 Outline of the thesis

The next chapter discusses the fundamentals of phishing attacks and some of the learning science principles that informed the PhishGuru design. Chapter 3 discusses relevant trust models, general aspects of security education, and relevant warning science literature. Chapter 4 introduces a trust model that we developed for the phishing scenario; we also present results from a study where we evaluated this model. Chapter 5 discusses the rationale for the embedded training concept and the evolution of the PhishGuru interventions. Chapter 6 discusses the motivation, study setup, and results from two laboratory studies that we conducted to evaluate the effectiveness of the PhishGuru. Chapter 7 presents the results of two real world studies conducted to test the effectiveness of the PhishGuru. Chapter 8 discusses two other phishing education systems that we designed, developed, and implemented. Finally, Chapter 9 presents conclusions from this thesis work and offers recommendations for security education.

Chapter 2

Background

In this chapter, in Section 2.1, we discuss some basics aspects of social engineering attacks. In Section 2.2, we discuss how phishing works, different types of phishing, and the life cycle of a typical phishing attack. We also discuss different countermeasures for phishing attacks. In Section 2.3, we discuss some relevant literature from the field of learning science.

2.1 Social engineering

A great deal of resources is spent on developing technical solutions that enable people and organizations to perform trustworthy interactions. However, these solutions are not foolproof. Kevin Mitnick writes "You could spend a fortune on technology and your network could still remain vulnerable to old-fashioned [social engineering] manipulation" [139]. Increasingly, social engineering attacks are being used to bypass highly secure systems. Social engineering is generally an act of con artists who manipulate the natural human tendency to trust. Social engineering attacks have existed for many centuries, but recently, various technologies have made it easier for criminals to conduct these attacks. Social engineers looking to conduct an attack use different sources to collect information about individuals or organizations: phone calls, dumpster diving, phishing emails [82]. Organizations lose lots of money and resources due to social engineering attacks. Studies have shown that it is easy to social engineer an employee of an organization [199].

Psychologists have studied why and how people can be persuaded. Some of the main factors involved in persuasion are: scarcity, consistency, social validation, linking, authority and reciprocation. People may fall for social engineering attacks because they are careless (e.g. opening an email from a bank with whom they don't have an account), because they want to be helpful (e.g. sending their bank account details to transfer money for a widow from Nigeria), because they feel it is in their comfort zone (e.g. a secretary at the front desk of an organization giving away some confidential information to a social engineer claiming to be the assistant to the CEO working from home) or because they are fearful (e.g. responding to an email requesting them to update their network access password in 24 hours or lose access to the network) [42, 118].

To combat social engineering attacks, strategies like strengthening security policies and increasing physical security have been proposed. But since social engineers take advantage of the persuasive nature of human beings, more effective measures include educating users about these attacks and providing resources to report these attacks [7, 83, 87, 192].

2.2 Phishing

Phishing is "a broadly launched social engineering attack in which an electronic identity is misrepresented in an attempt to trick individuals into revealing personal credentials (financial information, social security numbers, system access information and other personal confidential information) that can be used fraudulently against them" [66]. Victims get conned by spoofed emails and fraudulent websites that masquerade as a legitimate organization [93, 96, 116]. Broadly, phishing attacks can be classified into three different types: deceptive attacks, malware-based attacks, and DNS-based attacks. Deceptive attacks trick victims into giving their personal confidential information to spoofed websites. Currently, this is the most prevalent type of Internet attack. Malware-based attacks use phishing emails and websites to infiltrate a user's machine, where they execute malicious software. Keyloggers, session hijackers and web Trojans fall into this category of attacks. DNS-based attacks tamper with the integrity of the lookup process for a domain name. Content-injection, man-in-the-middle, and search engine phishing belong to this category of attacks [60]. Phishing has evolved over time, with phishers spreading across all sectors of business; however, the financial sector has been most affected [158].

2.2.1 Why phishing works

Users fall for phishing because of the poor online trust decisions they make. Psychologists have shown that people do not consider options when they make decisions under stress (i.e. accessing emails while busy at work). Studies have shown that people under stress tend to make decisions that are not rational and without thinking of all possible solutions [101]. Psychologists have termed this tendency the *singular evaluation approach*. In this approach, people tend to evaluate the solution options individually rather than comparing them to other options, ultimately selecting the first solution that works [104]. In *Human Error*, James Reason established that people use patterns and context to make decisions rather than looking at the analytical solution to the problem [159]. Generally, it is believed that people do not ask the right questions when making a decision. People are also primed by visible similarities and past experiences when making a decision [193, 194]. In particular, research has shown that non-experts make decisions without much thought, choosing the most obvious solution and the least strenuous path. Experts, however, make better decisions by thinking about many strategies [105]. These reasons (singular evaluation approach, failing to ask the right questions, and looking for patterns) help explain why people fall for semantic attacks.

In a laboratory study, Dhamija et al. found that 90% of a group of 22 participants were deceived by sophisticated phishing websites. Twenty three percent of the participants did not look at the browser cues such as the address bar, status bar, or security indicators. Dhamija et al. also categorized the reasons why people fell for phishing attacks: lack of knowledge (computer system, security and security indicators), visual deception (images masking text, images mimicking windows), and bounded attention (lack of attention to the presence or absence of security indicators) [53].

In a two-part study, Downs et al. analyzed the reasons why people fall for phishing attacks. In a laboratory study with 20 participants, they showed that an awareness of risks is not directly linked to useful strategies for identifying phishing emails. They also showed that participants who were able to identify familiar phishing emails correctly did not generalize their knowledge to unfamiliar phishing attacks [54]. In another online survey study among 232 participants, they showed that understanding how to parse URLs correctly and knowing the significance of the lock icon in the browser reduces vulnerability to phishing attacks. They also showed that knowledge about the consequences of phishing attacks does not affect behavior. Using these results, Downs et al. suggested that people be trained to understand and comprehend the cues rather than just being presented with warning messages [55]. Researchers have also shown that phishing attacks that use information from social networks and other contextual information are more effective than generic phishing attacks [90].

2.2.2 Life cycle of phishing attacks

Phishing attacks involve six phases, from planning the attack to removing all evidence of the attack [66]. We briefly describe each of these phases below.

1. *Planning*: In this phase, the phisher identifies the organization to spoof, decides what type of personal information to collect, and develops a story line or plot to use to collect the personal information. In this phase, phishers also select the technical infrastructure they will use to deploy the attack.

2. *Setup*: The phisher then designs attack materials such as phishing emails (*The lure*) and websites (*The hook*) [93]. Figure 2.1 shows an example of a phishing email spoofing the organization Citibank. The 'sender' address is masqueraded to look as though it comes from 'citibank.com,' but in reality it does not. The email also shows a sense of urgency and includes an action to be taken by the user. The link in the email is disguised to take the user to a

phishing website and not to the legitimate Citibank website. Phishers generally use open relays or zombie machines to send out the phishing emails.

3. *Attack*: Phishers use many vectors to perform the attacks. Some of these vectors are: websites, emails, instant messages, auto phone dialers (vishing) [156], chat rooms, blogs, bulletin boards, wireless networks, malware [66], search engines [60] and social networking websites. The most commonly used threat vectors are emails and websites [96].

   In this phase, phishers send out phishing emails to victims. Email addresses are harvested and collected from various sources; phishers then send out emails to these harvested email addresses. These email addresses are traded and reused among different groups of phishers. Phishers rely on users to click on the link in the email and go to the spoofed website to give their personal information. Figure 2.2 shows an example of a phishing website; phishers use such a website to collect personal information from users. The example shows the 'phishy URL' and the 'spoofed status bar.'

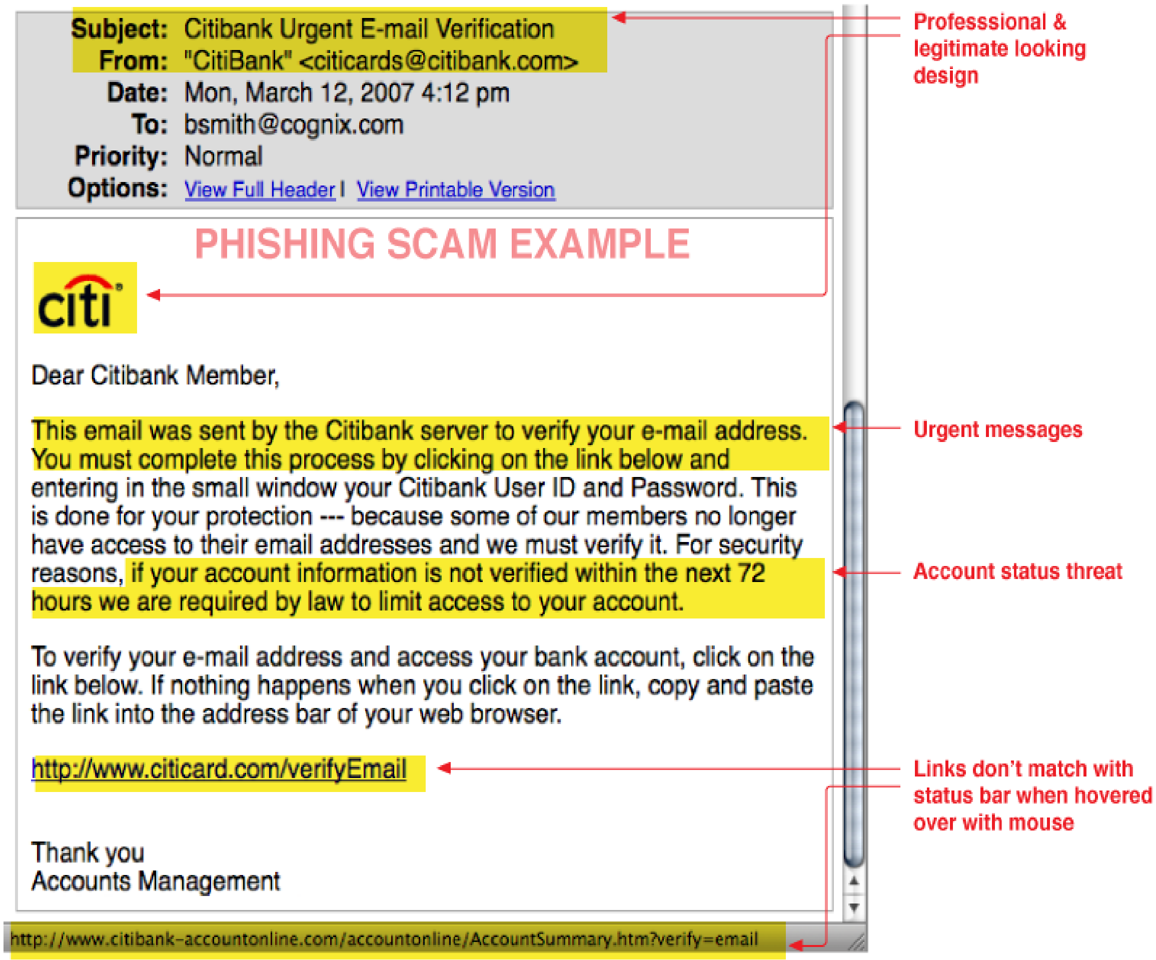
4. *Collection*: In this phase, phishers collect the personal information that victims provide to the phishing website [93]. The personal information that phishers tend to collect includes credit card numbers, social security numbers, computer and account login information, addresses for communication, and other sensitive pieces of personal information [204]. The personal information that users enter on the phishing website is either saved in files for the phisher to collect or is sent as an email to the phishers [116].

5. *Fraud & abuse*: Phishers sell, trade, or directly use the personal information collected from victims. Phishers employ *cashers* or *mules* to convert this information into cash or to engage in identity theft and other forms of fraud. Most of the time, mules are innocent people who perform this conversion without knowing that they are taking part in an illegal activity [93].

6. *Post attack*: During the post attack phase, phishers tend to remove all trails of their activity, including the phishing websites registered for the attack. It is believed that phishers also track the success of their attacks and use the knowledge they gain for future attacks.

2.2.3 Changing landscape of phishing attacks

The landscape of phishing has changed quite a bit since the days when phishers stole AOL account information. Back then, phishers sent out emails or AOL instant messages to potential victims asking for their AOL username and password. Later, phishers started sending emails like the example in Figure 2.1, assuming that some of the recipients would have a relationship with the spoofed organization. These phishing attacks can be considered *generic phishing.* The percentage of people who fell for these generic attacks was very low. Very soon phishers learned that they could



Subject: Citibank Urgent E-mail Verification
From: "CitiBank" <citicards@citibank.com>
Date: Mon, March 12, 2007 4:12 pm
To: bsmith@cognix.com
Priority: Normal
Options: View Full Header | View Printable Version

PHISHING SCAM EXAMPLE

Dear Citibank Member,

This email was sent by the Citibank server to verify your e-mail address. You must complete this process by clicking on the link below and entering in the small window your Citibank User ID and Password. This is done for your protection --- because some of our members no longer have access to their email addresses and we must verify it. For security reasons, if your account information is not verified within the next 72 hours we are required by law to limit access to your account.

To verify your e-mail address and access your bank account, click on the link below. If nothing happens when you click on the link, copy and paste the link into the address bar of your web browser.

http://www.citicard.com/verifyEmail

Thank you
Accounts Management

http://www.citibank-accountonline.com/accountonline/AccountSummary.htm?verify=email

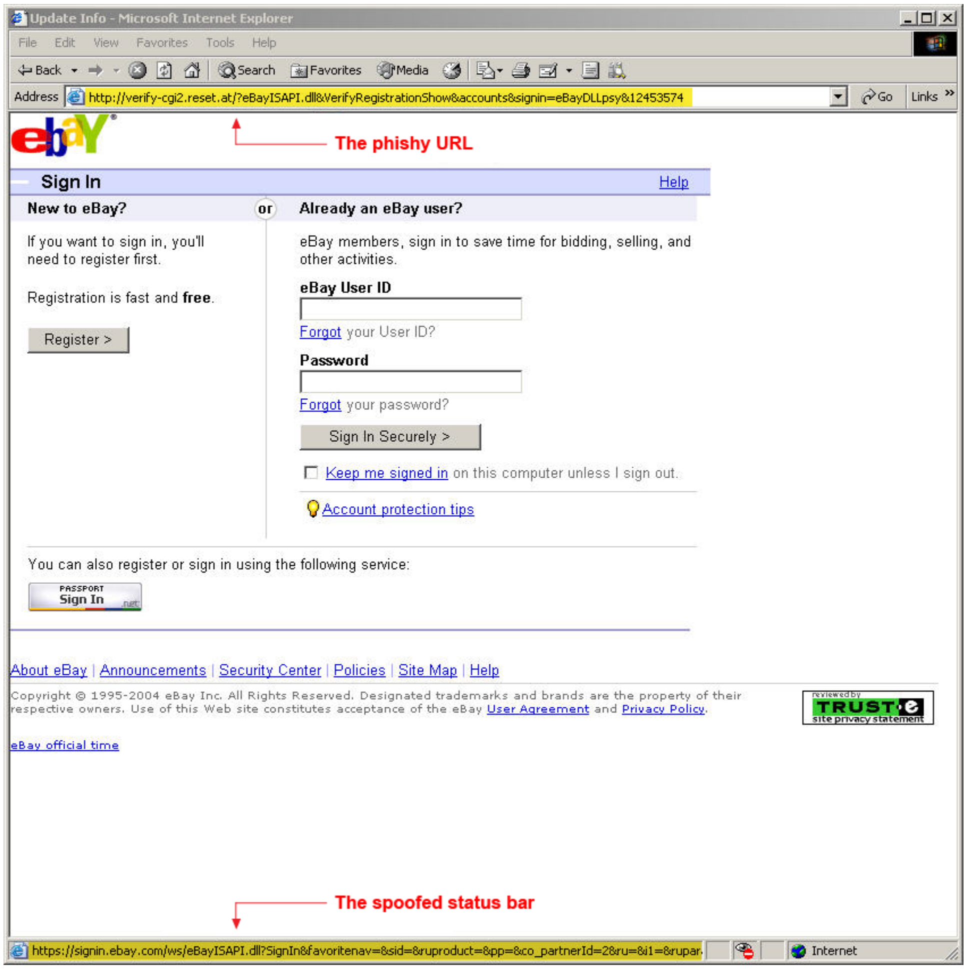
Professsional & legitimate looking design

Urgent messages

Account status threat

Links don't match with status bar when hovered over with mouse

Figure 2.1: Example of a phishing email spoofing the organization Citibank. Highlights the important characteristics of a phishing email.



http://verify-cgi2.reset.at/?eBayISAPI.dll&VerifyRegistrationShow&accounts&signin=eBayDLLpsy&12453574

The phishy URL

Sign In Help

New to eBay? or Already an eBay user?

If you want to sign in, you'll need to register first.

Registration is fast and **free**.

Register >

eBay members, sign in to save time for bidding, selling, and other activities.

eBay User ID

Forgot your User ID?

Password

Forgot your password?

Sign In Securely >

□ Keep me signed in on this computer unless I sign out.

Account protection tips

You can also register or sign in using the following service:

PASSPORT Sign In .net

About eBay | Announcements | Security Center | Policies | Site Map | Help

Copyright © 1995-2004 eBay Inc. All Rights Reserved. Designated trademarks and brands are the property of their respective owners. Use of this Web site constitutes acceptance of the eBay User Agreement and Privacy Policy.

eBay official time

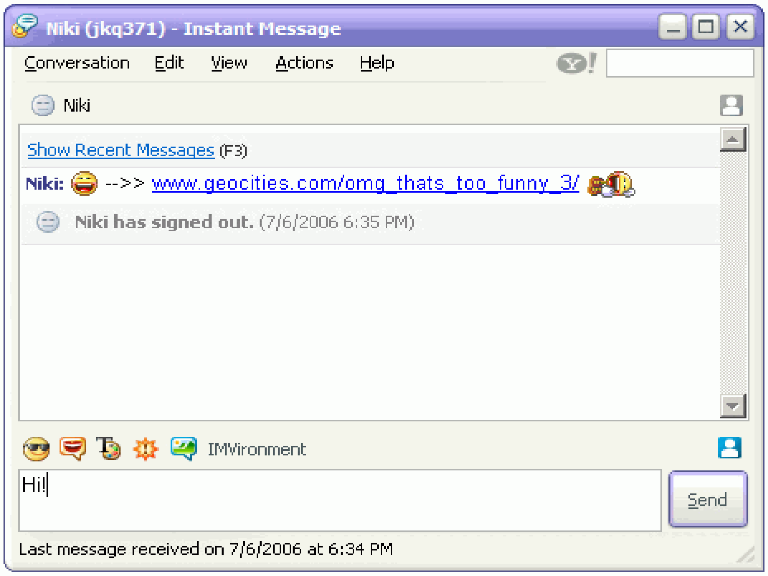
The spoofed status bar

https://signin.ebay.com/ws/eBayISAPI.dll?SignIn&favoritenav=&sid=&ruproduct=&pp=&co_partnerId=2&ru=&i1=&rupar.

Figure 2.2: Example of a phishing website spoofing the organization eBay. Highlights the phishy URL and the spoofed status bar.

use the large amount of personal information available on the Internet or through other sources to craft the phishing emails and increase the percentage of people who fell for the phishing attacks. Phishers started using their victim's personal information in the email (e.g. first and last name), a type of attack called *spear phishing* [136]. To further improve their chance of success and to increase the return on their investment, phishers started sending customized emails to executives and managers. These emails appeared to be official subpoenas from the United States District Courts or Better Business Bureau alerting them of a complaint against their organization. Some of these phishing emails pretended to come from a recruitment company or contractor seeking information about an invoice [73, 123]. This type of phishing attack is called *Whaling*. Recently, phishers have also started using social networking websites such as Facebook and MySpace to promote phishing websites [135]. Phishers also make use of the political situation to craft phishing emails. For example, many phishing emails and websites revolved around the 2008 US presidential campaign [94, Chapter 10].

In the last couple of years, phishers have also started to use Instant Messaging (IM) as an attack vector. They use all types of IM (e.g. Yahoo, Skype) to send the phishing message. For example, Figure 2.3 shows a link that is sent in Yahoo messenger. IM phishing is one of the SANS Top 20 security vulnerabilities for the year 2007 [171].



Niki: -->> www.geocities.com/omg_thats_too_funny_3/

Niki has signed out. (7/6/2006 6:35 PM)

Hi!

Last message received on 7/6/2006 at 6:34 PM

Figure 2.3: Example of phishing through instant messaging; here a user is getting a link to geocities.com.

Phishers also use other vectors such as Voice over IP (VoIP) to perform phishing attacks. Phishers target an area code and use VoIP to call numbers randomly informing potential victims that their credit card has been breached and asking them to call a specific phone number immediately. The numbers that these phishers provide are fake numbers which, when called, land in phishers' spoofed voice message systems. Phishers copy the exact voice message system from the actual credit card bank. When victims call the fake number, they are asked to enter their 16 digit card number and personal identification number (PIN) [156, 195]. These attacks are called *vishing*.

Phishers are now trying to steal information using a new technique from a tool kit created by the "Rock Phish" gang.[1] With this attack, phishers compromise a machine and use it to run many banks' websites. Phishers register meaningless domain names such as recy248.com and create URLs that look legitimate but have a randomly generated alpha-numeric character in the URL. These unique URLs are placed in the emails sent to potential victims. Existing filters don't have the capability to catch phishing emails with unique URLs. By using such advanced techniques, these phishing websites stay alive three times as long as the normal phishing website. Researchers have also found that sites derived from the Rock Phish kit account for more than half of the phishing attacks circulating on the Internet [142].

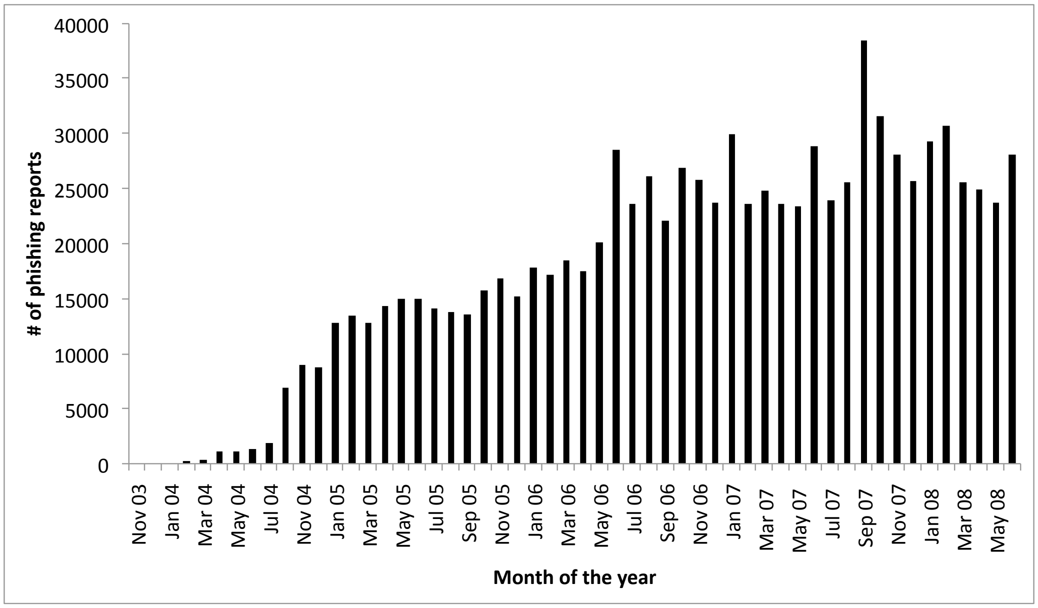


Figure 2.4: Unique phishing reports to Anti-Phishing Working Group (APWG).

A more sophisticated technique involves multiple nodes within the network (mostly compromised machines or botnets) registering and de-registering their address as part of the DNS NS record list. This process constantly changes the destination address for the addresses in the DNS zone. It therefore becomes difficult to identify the exact machine where the phishing websites are hosted. This technique is called "fast-flux"; it is believed that recent phishing attacks have made use of this technique [141, 184].

[1] These phishing attacks are mainly done by a gang who creates kits that even newbies or people with low technical expertise can use to conduct a very sophisticated phishing attack. This shows that, while the sophistication of the attacks has been increasing, the knowledge the phisher requires has been decreasing [96].

Figure 2.4 shows the increasing number of phishing reports that the Anti-Phishing Working Group receives every month. This includes all types of email phishing attacks discussed in this section.

2.2.4 Countermeasures for phishing

To protect people from phishing, a variety of strategies have been proposed and implemented. These strategies fall into three major categories: silently eliminating the threat, warning users about the threat, and training users not to fall for attacks. These categories of anti-phishing strategy mirror the three high-level approaches to usable security discussed in the literature: build systems that "just work" without requiring intervention on the part of users, make security intuitive and easy to use, and teach people how to perform security-critical functions [50].

Silently eliminating the threat

This strategy provides protection without requiring any awareness or action on the part of users. It includes finding phishing websites and shutting them down (law enforcement and policy solution) as well as automatically detecting and deleting phishing emails [65, 185]. Other methods that fall into this strategy are: DomainKeys from Yahoo!, which verifies the DNS domain of an email sender and the message integrity [207]; Sender Policy Framework, which uses Simple Mail Transfer Protocol (SMTP) to reject the forged address in the SMTP MAIL FROM address [179]; and Remote-Harm Detection (RHD), which collects the clients' Internet browsing history to identify phishing websites at the server end [92]. If phishing threats could be completely eliminated using these methods, there would be no need for other protection strategies. However, existing tools are unable to detect phishing emails with one hundred percent accuracy, and phishing websites stay online long enough to snare unsuspecting victims. Researchers have estimated that the mean lifetime of a typical phishing website is 61.7 hours, while rock phish domains are live for 94.7 hours. They also found that, for every day that one of these phishing sites is up and running, 18 users will be victimized [140, 142]

Warning users about the threat

A number of tools have been developed to warn users that the website they are visiting is probably fraudulent; these tools provide explicit warnings or interfaces that help people notice that they may be on a phishing website. Ye and Smith [208] and Dhamija and Tygar [52] have developed prototype "trusted paths" for the Mozilla web browser, tools designed to help users verify that their browser has made a secure connection to a trusted website. More common are web browser toolbars that provide extra cues; for example, a red or green light that informs users that they may be at risk by indicating the overall safety of the site [1, 150, 187, 188]. However, there are three weaknesses with this approach. First, it requires people to install special software (although newer versions of web browsers have such software included). Second, user studies have shown that users often do not understand or act on the cues provided by toolbars [138, 205]. Third, a study showed that some anti-phishing toolbars are not very accurate; even the best toolbars may miss over 20% of phishing websites [209].

Training users not to fall for attacks

There are many approaches to train users about phishing. The most basic approach is to post articles about phishing on websites; this approach has been utilized by government organizations [62, 63], non-profits [15] and businesses [56, 137]. Unfortunately, these articles fail to enhance user learning because of the difficulty involved in getting a large number of users to read these articles. A more interactive approach is to provide web-based tests that allow users to assess their own knowledge of phishing. For example, Mail Frontier has set up a website containing screenshots of potential phishing emails [120]. Users are scored based on how well they can identify which emails are legitimate and which are not. This approach has been applied mainly to test users rather than train them. Phishing education can also be conducted in a classroom setting, as has been done by Robila and Ragucci [164]. However, it is difficult to train large numbers of users through classroom sessions.

Another way to educate users is to send fake phishing emails in the interest of testing their vulnerability to these emails. Typically, at the end of such studies, all users are given additional materials to teach them about phishing attacks. This approach has been used with Indiana University students [90], West Point cadets [64], and New York state office employees [151]. The West Point and the New York state researchers conducted the study in two phases. In the first phase, participants did not have any prior preparation or training about phishing before being tested for their ability to detect phishing attacks. In the second phase, participants were given training materials and lectures about phishing before being tested again. Both studies showed that education led to an improvement in the participants' ability to identify phishing emails. Researchers have also started looking at non-traditional media, such as comic strips, to educate users about security attacks [91].

2.3 Learning science

Learning science examines how people gain knowledge and learn new skills. Very little formal work has been done to connect learning science literature to user education in the context of security. In this section, we will discuss some relevant learning science literature that we will later connect to end-user security education.

According to Clark, there are five types of content that can be learned: *facts, concepts, procedures,* *processes*, and *principles* [43]. People use these types of content to develop cognitive skills which help them process information and apply existing knowledge to a problem. Cognitive skills are developed by actively processing content types in the memory system. Human memory uses visual and auditory channels to process information and develop knowledge [44]. According to the ACT-R (Adaptive Control of Thought–Rational) theory of cognition and learning, knowledge can be classified as *declarative knowledge (knowing-that)* and *procedural knowledge (knowing-how)* [10]. Declarative knowledge is "factual knowledge that people can report or describe": for example, the Internet is a publicly accessible network of interconnected computer networks that transmits data using Internet Protocol. Procedural knowledge is "the knowledge of how to perform a task": for example, the steps involved in going through the email inbox and checking for new emails [10]. The ACT-R theory models procedural knowledge as a series of *production rules*, *if-then* or *condition-action* pairs [10]. In this thesis, we are interested in training users to create the right production rules for making online trust decisions. In general, existing educational solutions assume that users have declarative knowledge before they use the system [12]. However, since users do not have prior knowledge of how to protect themselves from semantic attacks, we will give them the declarative and procedural knowledge they need to avoid being victims.

The field of learning science has developed principles based on the way humans process information to acquire new skills and gain knowledge. In particular, learning science has developed instructional design principles that help users learn the content provided in training materials. Research has shown that when systems apply these design principles, they enhance learning [5, 12].

2.3.1 Instructional design principles

Researchers have developed instructional design principles that effectively educate users. In this sub-section, we will discuss some of the instructional design principles relevant to the work discussed in this thesis. Table 2.1 summarizes the instructional design principles that we used in training materials. We selected these principles because they are the most powerful of the basic instructional design principles and because they can be applied to online security training.

1. *Learning-by-doing principle*: One of the fundamental hypotheses of the ACT-R theory of cognition and learning is that knowledge and skills are acquired and strengthened through practice (by doing) [10]. Experiments in the cognitive tutor domain have shown that students who practice perform better than students who do not. In addition, students better understand instruction materials when they have to explain processes as part of their practice [6]. Research has also shown that difficult and random practice sessions are necessary for effective transfer of learning [57, 174]. In his book *Learning by Doing*, Clark Aldrich suggests simulations and games as ways to make people learn and practice better [4]. Learning by

doing is also being tried in traditional educational systems, where courses are being taught using a Story-Centered Curriculum (SCC) [172].

2. *Immediate feedback principle*: Researchers have shown that when tutors provide immediate feedback during the knowledge acquisition phase, students learn effectively, move towards more correct behaviors, and engage in less unproductive floundering [125, 174]. One of the principles developed by Anderson et al. in the context of tutoring is to "provide immediate feedback on errors [12]." Using LISP tutors, Corbett et al. showed that students who received immediate feedback performed significantly better than students who received delayed feedback [47]. Anderson et al. have emphasized that feedback should be immediate; otherwise, students begin to think about something else [11]. Research has also shown that simple forms of feedback, like "yes" or "no," and more detailed forms of feedback, like "the shot was off target to the right by 22mm," both encourage effective learning [57].

3. *Conceptual-procedural principle*: A concept is a mental representation of objects or ideas for which multiple specific examples exist (e.g. phishing) [43, Chapter 4]. A procedure is a series of clearly defined steps that results in the achievement of a given task (e.g. logging onto a computer) [43, Chapter 3]. The conceptual-procedural principle states that "conceptual and procedural knowledge influence one another in mutually supportive ways and build in an iterative process." [99] When learners encounter conceptual materials, they have to use a great deal of working memory, making the process very difficult. Presenting procedural materials in between conceptual materials helps reiterate the learned concepts. In this way, concepts reinforce procedures, and vice versa. Learners must understand both conceptual and procedural knowledge in order to develop competence in a given area [46]. This principle can be used to improve learning by providing conceptual and procedural knowledge iteratively. In an experiment where students learned about decimal places, students given concepts and procedures iteratively performed better than those who received them consecutively [99, 106].

4. *Contiguity principle*: Mayer et al. set forth the contiguity principle, which states that "the effectiveness of the computer aided instruction increases when words and pictures are presented contiguously (rather than isolated from one another) in time and space" [130]. Psychologists believe that humans make sense out of presented content by creating meaningful relations between words and pictures [44, Chapter 4]. In an experiment, students who learned about the process of lightning better understood the materials when words and pictures were close to each other (*spatial-contiguity*) [143]. In another experiment, students were asked to read a passage about vehicle braking systems. Students who received passages with words and pictures together explained the braking system better than a group in which words and pictures were presented separately [128]. In another experiment, Butcher et al. found that students who had visual and verbal information presented together performed better and had deeper

transfer than students who encountered the information separately [32].

Table 2.1: Instructional design principles used in this thesis.

Principle	**Explanation**
Learning-by-doing	People learn better when they practice the skills they are learning
Immediate feedback	Providing immediate feedback during the knowledge acquisition phase results in efficient learning
Conceptual-procedural	Conceptual and procedural knowledge influence one another in mutually supportive ways and build in an iterative process
Contiguity	Presenting words and pictures contiguously (rather than in isolation) enhances learning
Personalization	Using a conversational style rather than formal style enhances learning
Story-based agent environment	Using characters in a story enhances learning
Reflection	Presenting opportunities for learners to reflect on the new knowledge they have learned enhances learning

5. *Personalization principle*: This principle states that "using conversational style rather than formal style enhances learning" [44, Chapter 8]. People make efforts to understand instructional material if it is presented in a way that makes them feel that they are in a conversation rather than just passively receiving the information. To enhance learning, instructional materials should use words like "I," "we," "me," "my," "you," and "your" [129, Chapter 8]. In an experiment aimed at teaching arithmetical order-of-operation rules, students who received conversational-style messages were more engaged and, as a result, learned more than students in a control group [48]. Another experiment compared formal and conversational modes of presentation in a computer-based lesson on lightning formation. Results showed that students learned better when the information was presented in a conversational style [144].

6. *Story-based agent environment principle*: Agents are characters who help guide users through the learning process. These characters can be represented visually or verbally, and can be cartoon-like or real life characters. The story-based agent environment principle states that "using agents in a story-based content enhances user learning [145]." People tend to put more effort into understanding materials if an agent guides them through the learning process. Learning is further enhanced if the materials are presented within the context of a story [129, Chapter 8]. People learn from stories because stories organize events in a meaningful framework and tend to stimulate the cognitive process of the reader [104, Chapter 11]. Herman, for example, is an agent in story-based computer-aided lessons that teach users how

to design roots, stems, and leaves capable of surviving in eight different environments. Experiments showed that students in the group which had Herman guiding them outperformed the unguided group in terms of learning and correct decision making [145]. Many such agents have been developed to help users learn better [121]. It has also been found that the presence of the agents influences learning, but their features (cartoon like or real life) have little effect [203, 206].

7. *Reflection principle*: Reflection is the process by which learners are made to stop and think about what they are learning. Studies have shown that learning increases if educational systems include opportunities for learners to reflect on the new knowledge they have learned [46].

2.3.2 Measurements of learning

Research suggests that real world training should: (1) help a learner acquire new knowledge; (2) enable the learner to perform learned skills in the long run; and (3) enable him/her to transfer the learning to related and altered tasks [174]. These requirements can be used as a framework to measure user learning. There are many ways these requirements can be measured; for the work discussed in this thesis, we operationalize the measurements as follows:

1. *Knowledge Acquisition* (KA) is the ability of people to process and extract knowledge from instructional materials. Users should be able to use the acquired knowledge to make a decision in a given situation [19, 122]. This is usually evaluated by testing acquired skills just after learning by asking learners to repeat or apply the knowledge they have gained.

2. *Knowledge Retention* (KR) is the ability to retain or recall the concepts and procedures given by the content types when tested under the same or similar situations after a time period δ from the time knowledge was acquired. Researchers have frequently debated the optimum δ to measure retention. We will use the time difference dimension (δ) to measure KR. This has been one of the most frequently used dimensions to measure KR [19, 166]. In this thesis, we will measure retention where δ is anything more than one day. To test retention, researchers have used different time periods ranging from 1 day to 20 days [19, 27, 126]. If testing is done within one day of training, it is considered more of a test for knowledge acquisition than knowledge retention. One way to move acquired knowledge from training to the long-term memory of the user is by frequent testing [44]. There is a large body of literature on how to quantify retention; researchers have also created retention functions to describe the behavior of human memory [166].

3. *Knowledge Transfer* (KT) is generally the ability to apply learned concepts in a new situation. However, the precise definition of transfer and the measurement of transfer is heavily debated

in learning science literature [28,177,183]. Researchers have developed a taxonomy to classify and identify different types of transfer stated in the literature [23,72]. Two types of transfer discussed in the literature are immediate (near) transfer and delayed (far) transfer [69, 134]. Researchers have emphasized that transferability of learning is of prime importance in training. For the purpose of this thesis, transfer is the ability to extend the learning in one instance of a phishing attack to another instance after a time period δ. As with retention, there is also considerable debate about the optimum δ to measure knowledge transfer.

These design principles and forms of measurements were applied in the design and evaluation of PhishGuru.

Chapter 3

Related Work

In this chapter, we discuss some work related to this thesis. In Section 3.1, we discuss various trust models that have been developed to model users' behavior. In Section 3.2, we discuss some general aspects of security education; in Section 3.3, we discuss relevant warning science literature.

3.1 Trust and trust models

The literature on trust is vast, and the part of it that focuses on online trust is growing. Perhaps because of the multitude of angles from which it can be studied, trust is a concept with many dimensions [16, 17, 49, 127, 155] — a dynamic phenomenon [165], for which stable definitions and boundaries are still hard to define [81, 163]. Mayer et al. define trust as the "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control the other party" [127]. Corritore et al., in the context of online interactions, refer to trust as "an attitude of confident expectation in an online situation of risk that one's vulnerabilities will not be exploited" [49]. Different fields of study tend to focus on different aspects of trust. In economics, the focus is on agents' reputations and their effect on transactions [33, 36, 76, 84, 148, 157, 161]. In marketing, the focus is on strategies to enhance consumer persuasion and trust building [37, 40, 68, 189]. In psychology, trust is studied as an interpersonal and group phenomenon [170, 178]. And in HCI, the focus is on the relationship between system design and system usability [49, 162, 163].

In HCI literature, researchers commonly refer to two parties in a trust interaction: the trustee and the trustor. A trustor is a trusting party — for example, an individual who buys something online. A trustee is the party being trusted — for example, an online merchant [127]. A trustor enters into an interaction with the trustee only if the trustor's trust level in the trustee is more than a threshold value [191]. This threshold value is contextual and subjective to each trustee. This research focuses on the way the trustee interprets and makes use of information coming from the trustor.

Two other related concepts are often discussed in this literature: "trust" and "trustworthiness." Trust is a phenomenon demonstrated by a trustor who is placing his or her trust in a trustee, whereas trustworthiness is the characteristic the trustee displays [49]. This thesis work focuses mainly on trust and not trustworthiness: we are specifically interested in what causes people to place and misplace trust and how to improve that process rather than convincing people to trust more.

While uncertainty and risk generate trust issues, several other factors — called antecedents (of trust) — can positively or negatively affect the trustor's trust level. Table 3.1 shows some of the trust models discussed in literature, highlighting the specific dependent variable studied in each model as well as its antecedents. All of the models described in the table focus on online trust decisions (with the exception of Mayer et al.'s model).

Mayer et al. show that the literature includes many definitions of trust; researchers from different fields have been unable to come to a consensus concerning relevant definitions and boundaries. Mayer et al. also distinguish between "taking risk" and "willingness to take risk," with trust not being "taking risk *per se*, but rather *willingness* to take risk." They have established that trust in an organization is not only a function of the perceived ability, benevolence, and integrity of the trustee, but also the trustor's propensity to trust [127]. Their research also shows that perceived benevolence increases over time as the relationship between the trustor and trustee develops. Mayer et al.'s antecedents have been used in many other models [14,25,163]. In particular, Gefen evaluated Mayer et al.'s model using a survey instrument, finding that trust and trustworthiness should not be regarded as a single construct [75].

Bhattacherjee, in a theoretical conceptualization of trust, proposes a scale to measure individual trust in online firms. Similar to Mayer et al., his scale uses integrity, benevolence and ability as the main antecedents of trust. However, he adds the antecedent of familiarity, which might have a milder effect on trust than the other three [25].

Lee et al. [115] propose and evaluate 16 hypotheses about consumers' trust in Internet shopping. They evaluate the most common antecedents in the trust literature: ability, integrity, and benevolence. Mcknight et al. create a trust typology using constructs collected from 65 articles and books. They identify 16 categories of trust related characteristics, finding that benevolence and integrity are the most prominent antecedents [133].

Jarvenpaa et al. show that customers' willingness to buy from an internet store is based on both their initial trust level and their perception of the store's size, reputation, and potential risk. Ultimately, perceived reputation has a stronger effect on willingness to buy than the perceived size of the Internet store [97, 98]. Ruyter et al. show that trust in an e-service is dependent on the per-

Table 3.1: Trust Antecedents (arranged in chronological order); all models except for Mayer et al. [127] were developed for Internet transactions.

Authors	Measure variable	Antecedent factors
Mayer et al. [127]	trust in organizations	integrity, benevolence, ability, trustor's propensity to trust
Jarvenpaa et al. [97, 98]	willingness to buy online	perceived size, perceived reputation, initial trust in the store, perceived risk
Lee et al. [114]	trust on online stores	comprehensive information, shared value, communication
Mcknight et al. [132]	e-commerce consumer actions	disposition to trust, trust in technology and the Internet (institution-based trust)
Tan et al. [191]	trust for e-commerce transactions	party trust and control trust
Ang et al. [14]	likelihood of an online purchase	ability to deliver, willingness to rectify, personal privacy
Ruyter et al. [167]	trust in e-service	perceived risk, relative advantage, organization reputation
Bhattacherjee et al. [25]	trust in online firms	integrity, benevolence, ability, familiarity with trustee, willingness to transact
Corritore et al. [49]	trust in a specific transactional or informational website	perceptual factors — credibility, ease of use, risk
Chellappa et al. [40]	trust in e-commerce websites	disposition to trust, prior knowledge or experience, information from others, trustee's reputation, trust in information technologies
Chellappa et al. [39]	trust in online store	perceived privacy, perceived security, reputation
Riegelsberger et al. [163]	trust in technology mediated interaction	temporal, social, institutional ability, benevolence, motivation

ceived risk, relative advantage, and organization's reputation [167]. Chellappa et al. investigate the relationship between trust in an online store, perceived privacy, perceived security, and reputation, showing that trust is positively related to perceived privacy and perceived security. In particular, they find that the effect of perceived security on trust is greater than the effect of perceived privacy. People typically associate online problems with security rather than privacy issues; there appears to be a lack of knowledge about how to differentiate between the two [39].

Further research by Lee et al. shows that trust is dependent on the information that customers have when making a purchase decision (comprehensive information), common beliefs about the policies of the trustee (shared values), and timely sharing of meaningful information between buyers and sellers (communication). Lee at al. also show that transaction costs are negatively correlated with trust, while customer loyalty is positively correlated with trust [114].

Tan et al. posit that trust problems can occur because of hidden information; this problem arises before the parties agree to transact (*ex ante*). Alternatively, hidden action occurs when the problem arises after the transaction has been completed (*ex post*). Tan et al. further show that the trustor's level of transaction trust is dependent on their trust in the other party (party trust), as well as their trust in the control mechanism for successful performance of the transaction (control trust). When applying their model, the authors consider activities such as electronic payment and cross-border electronic trade [191].

Riegelsberger et al. showed that technology can transform the effect of trust-warranting properties. They focus on three scenarios to justify their model: the transformation of trust through technology, e-commerce, and voice-enabled gaming environments [163].

Ang et al. discuss the relationship between trust and privacy in ecommerce. They note that online customers don't like to wait to receive or see apparel or high-value products that they have bought, but they accept delays in order to buy CDs and books. The authors suggest that this may be an effect of the need to know that the purchase was not a mistake. Hence, their model shows that the likelihood of an online purchase is dependent on the sellers' ability to deliver, their willingness to rectify problems, and the personal privacy they offer.

Apart from Ang et al. and Chellappa et al., very few models consider privacy an antecedent for evaluating trust [14]. However, studies have shown that when online users have trust issues, their privacy concerns also increase [24, 196, 197].

3.2 Security education

The International Organization for Standardization (ISO) and National Institute of Standards and Technology (NIST) security standards, which many companies are contractually obligated to follow, include security training as an important component of security compliance [88], [154]. These standards describe a three-level framework that includes awareness, training, and education. Security awareness activities are intended for all employees of a company and often include videos, newsletters, and posters. Training, however, is generally meant only for employees who are involved with IT systems, and is mainly intended to provide basic computer security knowledge. This training is delivered primarily through classroom lectures, e-learning materials, and workshops. Education intended for IT security specialists is usually delivered via seminars or reading groups [153]. Even though ISO differentiates between training and education, research literature in learning uses these terms interchangeably. In this thesis, we also use the terms training and education synonymously. This thesis offers new approaches to increasing security awareness and delivering effective education, specifically on the topic of phishing.

There are two schools of thought on user education for phishing and semantic attacks. The first school believes that education will not work because “dumb users” cause most security problems [45]; the second contends that education will work because a “human firewall [the brain]” is the greatest defense [86]. Here, we argue that technology alone cannot adequately solve the problem of phishing and semantic attacks; instead, technology should be complemented by user education.

Some security experts have concluded that user education is not a solution for phishing and security attacks because “[education] doesn’t work,” “[education] puts the burden on the wrong shoulders” [152] and “security user education is a myth” [80]. Since security is only a secondary goal for users, some researchers believe that user education cannot be a solution [61].

Researchers have also formally tested whether user education helps users make better decisions [9, 89]. However, these studies fail to confirm that user education will not work. One study conducted pre- and post- phishing IQ tests to evaluate the effectiveness of Federal Trade Commission (FTC) phishing training material. Results showed an increase in the false positive (identifying legitimate emails as phishing emails) among participants who read the training materials. Researchers attributed this behavior to increased concern rather than increased knowledge of how to identify phishing emails [9]. However, the training material used for this study does not provide any specific principles for identifying phishing emails or websites; it is designed for identity theft in general and not specifically for phishing. Therefore, these training materials only raise general awareness or concern for the phishing problem, rather than providing the knowledge needed to accurately identify phishing emails.

Another study evaluated the extended validation feature in Internet Explorer (IE) 7 [89]. This study measured the effects of extended validation certificates, which only appear on legitimate websites, and the effect of reading a help file about security features in IE7. The results showed that participants who did not know about extended validation did not notice the indicator. This study also found that participants who read the training materials on general security classified both legitimate and fake websites as legitimate when the warning did not appear. Based on these findings, the researchers claimed that training did not help users make better decisions. However, the training materials used in this study were designed to teach how extended validation works and how the indicator mechanism is used in IE7. The training materials did not provide any specific tips on how to identify phishing websites. Both of the above studies [9, 89] used training materials which had a broader purpose, like teaching people about identity theft or general security. Therefore, these results do not necessarily show that user training does not work in the context of phishing. In general, most of the online training materials increase user suspicion rather than providing specific training on how to identify phishing emails or websites. Hence, training materials have to be developed specifically for phishing and semantic attacks, materials that will help users make better online trust decisions.

Some researchers argue that training can help users avoid falling for phishing and security attacks [21, 26, 60, 79, 96, 102], [51], [93]. Some researchers also believe that educating the end user is one of the best lines of defense an organization can use to combat security attacks [86]. Organizations spend a considerable amount of money on security training for their employees [79]. Studies have shown that "the majority of users are security conscious, as long as they perceive the need for these [secure] behaviors [2]." Thus, user training is important in combating phishing and other semantic attacks.

Little research has been done on how to design instructional materials to educate people about phishing and semantic attacks. This thesis, however, focuses on the content and presentation of training materials designed to educate users to avoid falling for phishing attacks. We argue that automated detection systems should still be used as the first line of defense against phishing attacks; however, since these systems are unlikely to perform flawlessly, they should be complemented by education to help people better recognize fraudulent emails and websites. A better informed user will be the best defense against phishing attacks.

In this thesis, we focus on the design and evaluation of email interventions in order to understand what kinds of designs are most effective; that is, which designs teach people about phishing and actually protect them in practice. This work aims to teach people what cues to seek in order to make better decisions in more general cases. For example, rather than just teaching people to avoid PayPal phishing attacks, we want people to learn how to identify phishing attacks in general.

3.3 Warning science

Warnings are used to inform people of an impending or possible hazard, problem, or other unpleasant situation. The main purposes of warnings are: to communicate important safety information; to influence or modify people's behavior; to reduce or prevent health problems; and to serve as a reminder. In designing a usable system, designers should try to eliminate the hazard through good design. If they can't, they should at least guard against the hazards. Only when these two strategies don't work should designers fall back on the option of developing good warnings. Warnings should serve as a supplement for good design rather than a replacement [200]. Researchers have developed and evaluated guidelines for developing warnings. Among many other standards in designing warnings, the American National Standards Institute has a standard ANSI Z535.2 specifically for designing safety signs [8]. This standard recommends that warnings present the following information: what the hazard is, instructions on how to avoid the hazard, and the consequences of not avoiding the hazard. These recommendations have been supported by research [202]. Warnings have been applied in different fields for traffic signals, aircraft control, railroads, poison bottles, refrigerators, and cars. It is believed that well developed warnings can change human behavior.

Similar to aircraft and railroads, software systems such as email clients and web browsers also present warnings to protect users from hazards. Little research has been done on how to apply formal methods or frameworks to designing the warnings presented in software systems. Warnings used in software systems can be classified as *passive* or *active* indicators. Passive indicators are warnings presented to users somewhere in the system to help them make a decision. For example, web browsers and email clients alert users about phishing websites and spam emails, respectively. Some toolbars provide labels using a color scheme like red, yellow, and green for respectively, “bad,” “don’t know,” and “good” (or known) websites. Figure 3.1 shows one of the passive browser indicators from phishing toolbars. Active indicator warnings, alternatively, stop users’ mid-task, presenting them with information that helps them take action to avoid or mitigate the danger. Figure 3.2 shows an active warning message from the Firefox 3 web browser.

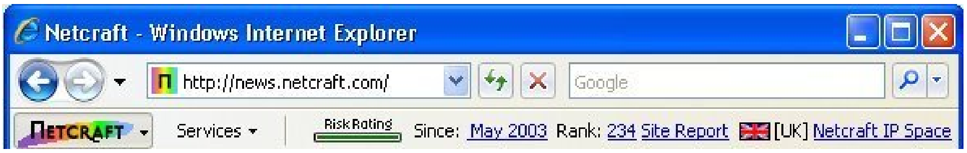
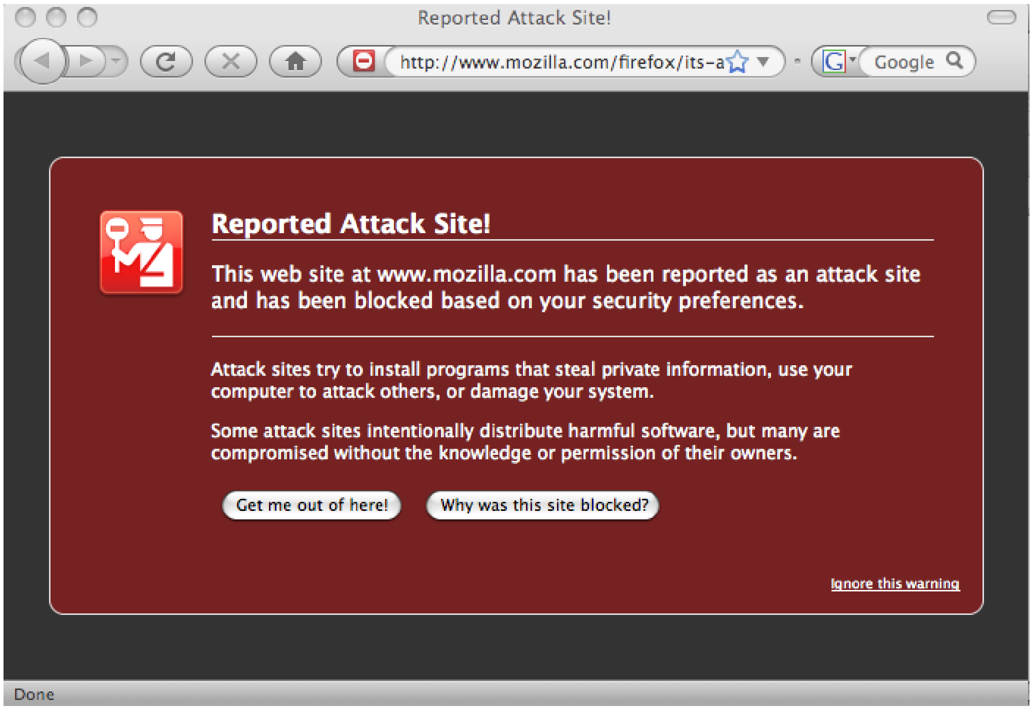


Figure 3.1: Passive warning in Netcraft. Presents information like risk rating of the website, year of registration of the website, popularity of the website among toolbar users, and an image of a flag or a two letter country code where the website is hosted.



Reported Attack Site!

This web site at www.mozilla.com has been reported as an attack site and has been blocked based on your security preferences.

Attack sites try to install programs that steal private information, use your computer to attack others, or damage your system.

Some attack sites intentionally distribute harmful software, but many are compromised without the knowledge or permission of their owners.

Figure 3.2: Active warning in Firefox. Presents only three choices which users can use to avoid or mitigate the danger.

Wu et al. showed that active phishing warning indicators are more effective than passive warnings displayed in the toolbars [204, 205]. They found that 25% of participants failed to notice the warnings or indicators presented by the toolbars at all, and that a large percentage of participants who noticed the toolbar indicators still presented information to the fake phishing websites. With a similar goal of comparing active and passive warnings, Egelman et al. in a laboratory study asked participants to buy a product on eBay and sent them a related eBay fake phishing email. Results from this study showed that 79% of the participants heeded the active warnings and closed the phishing websites, while only 13% of participants even saw the passive warnings. Based on these results, Egelman et al. suggest that phishing warnings should interrupt users' primary tasks, provide clear choices for users, and prevent habituation by making the phishing warnings different from other less serious warnings [58]. Researchers have also shown that using signal words in warnings gets the readers' attention and makes warnings more effective [85]. A few examples of signal words are *important*, *caution*, *unsafe*, *warning*, *danger*, and *critical.* Learning science principles discussed in Section 2.3.1 can also help make warnings more effective.

Since humans perform security-critical functions, like making a decision when a phishing warning is presented to them, Cranor developed a framework for reasoning about the human in the loop in any security systems [50]. This framework was based on Wogalter's model for Communication-Human Information Processing (C-HIP) [201]. Cranor's framework helps to enumerate the activities that systems expect humans to perform in a security-critical situation. This framework considers the characteristics of the information that is presented to the humans, as well as the personal characteristics and capabilities of the particular humans involved. This framework can be used to investigate the reasons why a particular communication is effective or ineffective. Using Cranor's framework to evaluate the effectiveness of the type of warning in Figure 3.2, one can ask the following questions: Is this the best type of warning message for this situation? What relevant knowledge or experience do the users have? Are users capable of taking appropriate action? Do users understand the warning message? Are users motivated to take appropriate action?

Masone applied Cranor's framework to evaluate his email system, a system called Attribute-Based, Usefully Secure Email (ABUSE). ABUSE is an email program designed to help users make informed trust decisions about emails they receive. Using Cranor's framework, Masone found that users' graphical user interface effectively allowed them to understand the attributes they needed to possess in order to make correct decisions while viewing emails [124].

Chapter 4

Expert and Non-expert Decision Making in Online Trust Scenarios

> This chapter is joint work with Alessandro Acquisti, and Lorrie Cranor. An earlier version of the content in this chapter was published at PST 2006 [107].

In this chapter, we will discuss both a trust model we developed and the results of a study we conducted to evaluate that model. In Section 4.1, we highlight the relationships between asymmetric information, trust, and expert modeling. In Section 4.2, we introduce the trust model and explain how it can be used to understand, represent, and contrast expert and non-expert decision processes. In Section 4.3, we describe the data collection protocol that we used during the interviews. In Section 4.4, we present the demographics of participants and the results of the study. Finally, in Section 4.5, we discuss some implications of the results.

4.1 Experimental approach: From signalling theory to expert modeling

Phishing attacks exploit the gap between what the system actually does (the system model) and what the user thinks the system is doing (the user model). This is a quintessential problem of misplaced trust, which – borrowing a term from economics literature – can be represented as a problem of *asymmetric information* [3, 190]. Asymmetric information refers to scenarios where each party to a transaction has unequal access to information about that transaction. For instance, the person who receives an email, ostensibly from her bank, has less information than the actual sender of the email about whether the email is legitimate. Such unequal access to information may disrupt transactions and cause consumer losses: consumers may refuse to follow the instructions in a legitimate email and incur costs from a false positive error; alternatively, they may respond to a scam message meant for identity theft (a false negative error) and incur even more significant costs.

The concept of a gulf between system models and user models echoes the signal theory of asymmetric information. In this theory, one party to a transaction sends a signal to other parties in order to alleviate inefficiencies created by the lack of reliable information. This signal contains information that the sending party considers relevant to the transaction [186]. However, the receiving party needs to evaluate whether the signal is an honest depiction of the underlying state of certain variables. Phishers try to make their emails as legitimate-looking as possible, twisting the system model to take advantage of the user model; sometimes they will even include warning messages about phishing threats in their attacks. End-users need to make judgments based on the signals available to them about the actual legitimacy of the email they have received.

A number of things can go wrong in this process, each leading to incorrect assessments and improper assignments of trust. For instance, psychologists have shown that when people are under stress (e.g., accessing emails while busy at work), they do not consider options when they make decisions; they tend to make decisions that are not rational, or ones that fail to consider various alternative solutions [101]. Psychologists have termed this the *singular evaluation approach* (see also [104]). In his book *Human Error*, James Reason posits that people use patterns and context to make decisions instead of looking at the analytical solution to the problem [159]. When making new decisions, individuals are also primed by visible similarities and their previous experiences [193, 194].

Indeed, researchers have investigated the similarities and differences between experts' and non-experts' decision-making processes across numerous fields in order to improve non-experts' skills and abilities: from chess [38, 105] to physics [41], from privacy and security [34] to risk communications [146]. Such research has shown that, in certain scenarios, experts developing their game strategy rely on a small set of possible actions, within which they find a satisfactory course of action [105]. Experts also tend to combine information from long-term memory with "chunks" (smaller pieces of information) from short-term memory [77], using multiple models of a domain to make a decision. Conversely, non-experts often categorize problems more simply than experts; as a result, their mental representation of a domain tends to differ [182]. In particular, research has shown that non-experts often choose the most obvious solution or least strenuous path, while experts in the same scenario would usually consider multiple strategies [105]. Experts, however, do not always solve a problem better than non-experts: Goldberg found that clinical judgments made by experts were not significantly better than those made by non-experts [78].

The distinction between expert and non-expert computer users is likely to be useful in the context of online trust [34,35]. In a related study of phishing, Downs et al. found that people who knew the correct definition of phishing were less likely to fall for phishing emails. They also found that people who correctly interpreted online images indicating site security were less likely to click on a phishing link or give personal information to phishing websites [55]. In the rest of this chapter, we extend that type of analysis. We combine the conceptual frameworks of asymmetric information and signalling (from economics), with a mental model interview methodology (from the expert literature). The former offers a lens through which we can analyze the dichotomy between system and user models, investigating the points of highest vulnerability in the decision-making process of users involved in phishing scenarios. The latter identifies how experts and non-experts handle decision making by using available signals. Combining a signalling approach with the expert modeling approach, we are able to understand opportunities and shortcomings in the trust-sensitive decision-making processes of Internet users.

4.2 A simple trust model informed by signalling theory

In this section, we represent trust decision problems as stylized combinations of ideal components and relationships, as shown in Figure 4.1.

States of the world are the true realizations of the variables that affect both a user's well-being, and secondarily, their decision-making process. Example states of the world are whether a certain email was actually sent by a colleague or by a spammer, or whether the Citibank page a user just accessed is a legitimate page residing on the bank's servers or a fraudulent site residing on a malicious third-party host.

Signals are pieces of information available to a user about the states of the world. They are, in other words, noisy functions that underlie the states of the world and may be more or less informative. Examples include the "from" field in an email, or the URL address at the top of a browser. Most of the literature on trust addresses signals – for example, credibility, ease of use [49], and benevolence [163].

Actions are the set of things a user may do in a certain scenario. For example, a user who receives an email that (ostensibly) appears to be from a colleague but contains a suspicious attachment may open it, delete it, scan it with anti-virus software, or ignore it.

Decisions refer to both the adoption of a specific set of actions and the strategy that governs that adoption. The decision-making strategy may be determined by personal heuristics or rational deliberation, and may be informed by the evidence available to a user about the states of the world and the consequences of her actions. Through her decisions, the user attempts to attain some objective measure of well-being. For example, a user may decide to delete an email when the sender is unknown.

Attackers (e.g., spammers or phishers) are entities that deliberately influence signals and states of the world — and therefore a user's decisions — to their own advantage.

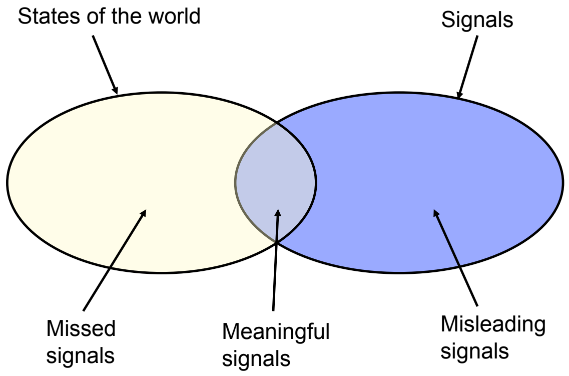


Figure 4.1: Generic trust model showing the model elements.

In other words, users make *decisions* among alternative *actions*, in order to satisfy certain personal well-being objectives. Such decisions are informed by noisy *signals* about the true underlying *states of the world*. External *attackers* can affect these world states, signals, decisions, and, ultimately, a user's well-being. We define online trust problems as those that arise when dichotomies between signals and underlying states affect the user's decisions.

Figure 4.1 focuses on a stylized relationship between states of the world, signals, and actions taken by the subjects. The various layers are used to represent actual states of the world and signals available to the consumer (the areas depicted in Figure 4.1 should be interpreted as distinct, overlapping layers rather than sets of a Venn diagram). The layer on the right represents the information available to and used by the consumer in her decision process – regardless of whether the signal is an accurate depiction of the underlying state of a corresponding variable, and regardless of whether or not the individual interprets the signal correctly. The left layer represents states that may affect the user's decision process.

The signals which lie at the overlap of the states of the world set and the signals set can be considered *meaningful signals*. For instance, an individual may use an email's subject and sender information as signals to decide whether to open or delete it; those signals are *possibly* quite informative about the identity of the true sender of the email, and therefore relate to states of the world that may be relevant to the user and ultimately affect her well-being.

However, certain pieces of information may affect a user's decision making process without being truly representative of the underlying states. As shown in Figure 4.1, these are called *misleading signals*. For instance, a user working to determine whether an email is legitimate may rely on an uninformative or misleading signal — e.g., the fact that the email is signed by a known, trustworthy bank.

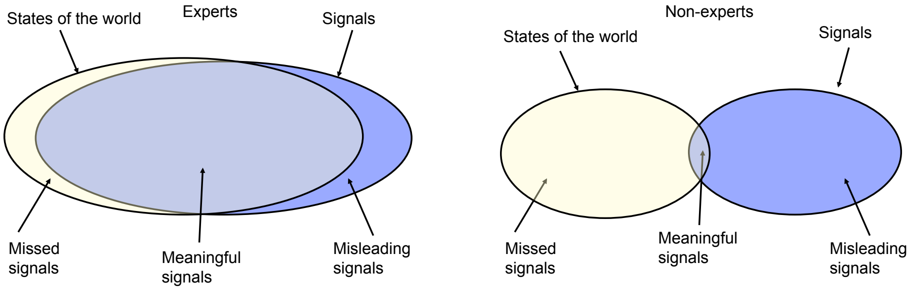


Figure 4.2: Left: In the expert model, it is hypothesized that meaningful signals are proportionally greater in number than misleading or missed signals. Right: In the non-expert model, it is hypothesized that misleading and missed signals are proportionally greater in number than meaningful signals.

In Figure 4.1, signals relating to underlying states of the world that may affect a user's well-being but are ignored or not considered are termed *missed signals*. For instance, a user may not be aware that envelope information in an email may show the route taken by that email and therefore reveal its actual sender.

One of the goals of this research is to populate the stylized model represented by Figure 4.1 with the signals actually used by experts and non-experts; this would serve to contrast the relative prominence of the various layers for different types of users. For instance, a number of signals meaningful to an expert may be completely missed by a non-expert, while a signal that a non-expert may consider meaningful could, in fact, be considered misleading by an expert. In other words, experts and non-experts may differ in their ability to detect and interpret signals and ultimately assess their association with the underlying states of the world. The experts' version of the representation in Figure 4.1 might look like the left side of Figure 4.2, in which one can detect a significant overlap between states and signals (experts make significant use of meaningful signals, with fewer missed or misleading signals than non-experts). The non-experts' version, however, may look like the right side of Figure 4.2, where we hypothesize a narrower overlap between states and signals. This, in turn, creates a small space for meaningful signals and a large space for missed and misleading signals. In comparison to the experts' model, the area which indicates missed and misleading signals is much larger.

4.3 Methodology

In the previous section, we introduced a stylized model of online trust decision making processes that was informed by signaling theory. The model is not meant to be predictive in the positive sense of the term: its utility lies in providing a framework to represent the different ways in which experts and non-experts make use of (or ignore) information available to them. In this section and the following one, we present the methodology and results of an application of this model to a phishing scenario, in which a user receives an email from what looks like a legitimate sender [*signal*]. The email may have actually been sent by a scammer [*attacker*], with the linked site the user visits turning out to be a phishing site [*state of the world*]. Users can choose whether or not to follow the email instructions [*actions*], proceeding based on personal heuristics, knowledge, and the expected consequences of their choices [*decision*].

In the rest of this section we highlight how we used interviews with experts and non-experts to understand their decision processes and what they knew about signals, actions, and attackers in an online scenario. We adopted a mental model interview approach [146] and populated the model with the qualitative data extracted from the interviews.

4.3.1 Screening and recruitment

We recruited interview participants by posting flyers in various locations around the city of Pittsburgh, sending messages to online mailing lists, and posting classified ads on Craigslist.org. These solicitations invited participants to apply for an interview study at Carnegie Mellon under the following qualification criteria: participants had to have an email account, be able to travel to Carnegie Mellon's campus, and be at least 18 years old. We invited the initial respondents to answer an online screening survey which contained three screening questions asking: 1) whether the participant had ever changed preferences or settings in their web browser; 2) whether the participant had ever created a web page; and 3) whether the participant had ever helped someone fix a computer problem. We classified as "non-experts" those who answered "no" to all three questions (the same approach and questions had been successfully used in other studies [54]).

Individuals doing research in the area of security and privacy at Carnegie Mellon University and the University of Pittsburgh were recruited as "experts." We recruited 14 non-experts and 11 experts for hour-long face-to-face interviews.

4.3.2 Interview Protocol

The interviews focused on three elements of the trust model presented above: the signals that the participants do or do not rely upon in their online decision making processes; the actions they consider taking; and their knowledge and awareness of attackers who may try to deceive users.

The interviews were conducted one-on-one in a closed laboratory, voice recorded, and transcribed for later analysis. The participants were told the following about the goals of the study:

> "The purpose of the study is to learn how people make decisions about what to trust

> and what not to trust. There are no right or wrong answers. We plan to use the information from the interview to assist in the development of tools, such as new e-mail client applications. If at any time you wish to terminate your participation or skip questions in this study, you have the right to do so without penalty. However, every response is important and we highly value your contribution towards the discussion."

After a few preliminary questions about the participant's usage of computers, operating systems, email clients, and web browsers, we began a series of questions related to the concept of "trust." The first question asked what applications (if any) the participant used to decide whether the organization they were interacting with on the Internet was trustworthy; later questions focused on what "trusting an email" or "trusting a website" meant for the participant. Specifically, we asked: "Could you please tell me what it means to you to 'trust' an email?" We found that both experts and non-experts were aware of potential trust issues online; experts were aware of *specific* issues, but non-experts were only aware that trust issues exist. A typical response from an expert was: "trust an email basically means that you trust where it comes from, as well as the integrity of the email, that it is what it says it's from." A typical response from a non-expert was: "it's just to make sure that it's not spam or that it contains any viruses or anything that will harm my computer." We found similar differences among experts and non-experts for the question "Could you please tell me what it means to you to 'trust' a website?"

After discussing the concept of "trust" with the subjects, we asked questions about trust-related imaginary scenarios. Specifically, we asked the participants to imagine facing seven different online scenarios. The scenarios are listed in Table 4.1. Four scenarios relate to emails and three relate to websites (the main vector for phishing attacks are emails; however, a successful phishing attack does not stop with the victim opening and reading the message, but goes on to require her to access a website and provide personal information). Three scenarios focus explicitly on phishing; four scenarios are indirectly related to phishing, but focus on online trust decisions.

For each scenario, we asked experts and non-experts to discuss what kind of things they would look for in order to decide whether or not to "trust" [the email, the website, etc] (we inquired about the "signals" the user would and would not rely upon, although we did not use that term during the interviews); what kind of things they would do if they actually faced such a scenario (the "actions"); and whether they believed that a third party might manipulate some of the aspects described in the scenario in order to influence the participant's decisions (the "attackers").

Specifically, to collect information about the signals that participants knowingly use in their decision-making process, we asked "What are the possible things that you might look for [in the scenario] to help you make your decision?" We then collected information about the perceived helpfulness of signals by asking "On a scale of 1–7, how much information does [the signal mentioned] give you to help make your decision about [the scenario], where 1 is not at all useful and 7 is very

Table 4.1: Scenarios described to the participants, presented in the same order they were discussed with participants.

Scenario	Description of the scenario provided to participants
General email	"Assume that you receive a new email message in your inbox..."
Email with account	"Assume that you receive an email from a bank that you have an account with, asking you to update your personal information using a link provided in the email, or else your account will be terminated..."
Email with no account	"Assume that you have an email from a bank that you do not have an account with, asking you to update your personal information using a link provided in the email or your account will be terminated..."
Email with attachment	"Assume that you get an email that has either a .zip file or an .exe file as an attachment...."
Auto download	"Assume that you are working on some personal stuff on the internet (e.g. looking up certain medical / financial information) and the website you are accessing prompts you to download software from another website to view the contents..."
Deliberate download	"Assume that you decided to download a software from a website, for example, a music player [could be any other software also]..."
Buying a book online	"Assume that you want to buy a book from a website for your personal use using your personal credit card..."

useful?" In order to understand the confidence level a user had in each signal in a given scenario, we asked "On a scale of 1–7, how confident are you that [the signal mentioned] will help you make a good decision about [the scenario], where 1 is not at all confident and 7 is very confident?" If the participant did not remember all the signals she had mentioned, we reminded them. These scores helped us determine which signals the subjects perceived as meaningful to their decision process.

To collect information about the different actions that the participants take, and the likelihood that they would perform such actions when facing the various scenarios, we first enquired "What are all the different things that you could do in this situation?" and then followed up that question by asking: "On a scale of 1–7, how likely do you think it is that each of the possible things mentioned above can happen [that you would do], where 1 is not at all likely and 7 is most likely?" If participants did not remember all the actions *they* had mentioned, we reminded them.

In order to understand experts' and non-experts' knowledge of the potential influence of external parties on their decision making process, we asked "Is there any third person who might change the things that you look for while making your decision? If so, who are they?" If the participants were not sure about the meaning of the question, we mentioned "these people can change your decision, e.g. somebody who can [do something relevant to the scenario].”

Because this study was based on hypothetical scenarios rather than monitoring actual behavior, we were comfortable with the fact that both experts and non-experts may have discussed items that they would not be likely to use in real life. This analysis focused on experts and non-experts’ awareness of and reliance on information in online trust scenarios. In related studies [109], we complement this mental models approach by analyzing behavior in controlled experiments.

4.3.3 Content coding and analysis

Interviews were transcribed verbatim, with responses coded for each of the model element questions. Each signal a participant mentioned was assigned a unique code, while the associated 7-point rating (usefulness of a signal or likelihood of a certain action) was coded with the signal. Similarly, we coded the actions mentioned by participants with their associated 7-point ratings (likelihood that participants mentioned they would engage in the action). We also coded attacker information along with information provided by the participants for all other questions.

Analyses were performed using R 2.7.0 and Microsoft Excel for Macintosh. We performed analyses on (1) dichotomous variables (e.g. whether participant mentioned a signal or not) and (2) categorical variables (e.g. the 7-point rating the participants provided for each signal). We also marked some of the most interesting responses in order to quote them in the following discussion.

4.4 Demographics and results

In this section, we present the analysis of the results. Specifically, after summarizing the demographics of the participants in Section 4.4.1, we discuss what the interviews taught us about how participants do or do not use signals in Section 4.4.2. Using the data we collected, we show that experts and non-experts use significantly different types of signals to make their decisions. In Section 4.4.3, we present the actions discussed by interview participants. Results suggest that subjects’ skill levels (expert or non-expert) affect which action they claim they would perform first; however, there is no strong relationship between the level of expertise and the overall set of actions mentioned by experts and non-experts. In Section 4.4.4, we discuss how experts and non-experts perceive online attackers. Notwithstanding media attention to these problems, results show that non-experts remain largely unaware of the ease with which a third party can spoof emails and websites.

4.4.1 Demographics

Participants reported computer experience ranging from less than two years to more than 16 years and Internet usage ranging from 6–15 hours per week to more than 51 hours per week. Expert participants had more years of computer and Internet usage than non-expert participants: while 45% of the expert participants had 11–15 years of computer experience, only 14% of the non-experts fell in this category. Also, 55% of the experts reported using the Internet in the range of 31–50 hours per week, while only 7% of the non-experts fell in this category. The average number of emails that the experts received per day (137) was significantly greater than the number received by non-experts (28). In addition, 91% of the participants in the experts group were graduate students, while only 57% were graduates in the non-experts group. Other statistics about the sample group are provided in Table 7.7. The expert interviews ranged in length from 54 minutes to 100 minutes (mean = 70.1, SD = 12.7), while the non-expert interviews ranged in length from 32 minutes to 60 minutes (mean = 46.7, SD = 8.9).

Table 4.2: Characteristics of participants.

	Experts	**Non-experts**
Characteristics	N = 11	N = 14
Gender		
Male	100%	36%
Female	0%	64%
Operating System		
Windows	73%	100%
Linux	18%	0%
Mac OS	9%	0%
Browser		
IE	55%	93%
Firefox	36%	7%
Safari	9%	0%
Email client/service		
Microsoft Outlook	46%	7%
Yahoo	0%	29%
Others	56%	64%
Occupation		
Students	91%	57%
Others	9%	43%
Avg. emails per day	137	28
Avg. age	26	32

All non-experts used Microsoft Windows, while 73% of experts used Windows, 18% used Linux, and 9% used Mac OS. Forty-six percent of the experts used Outlook as their email client, while only 7% of the non-experts used Outlook. Other email clients used by experts included Gmail, Hotmail, Evolution, Thunderbird, and Mail.app. Other email clients used by non-experts included Gmail, Hotmail, Mulberry, and AOL. When we asked participants whether there were any applications that they used to help decide whether the emails they receive were trustworthy, 36% of the experts mentioned Pretty Good Privacy (PGP) encryption tools, while none of the non-experts mentioned any tools.

4.4.2 Signals

Signals refer to the information that participants claimed they would use when making decisions in the scenarios we presented. We asked the interview participants questions such as "What are the possible things that you might look for [in an email] to help you make your decision?" for all the scenarios that were related to email. We recorded the number of signals mentioned in each scenario and by each participant. During the interview, the discussion about signals was linked to the discussion about the various actions the subject could take when facing a given scenario – such as opening the email, reading it, forwarding or deleting it, and so forth (see Section 4.4.3 below).

We found significant differences between experts and non-experts in terms of the type and number of signals mentioned. Column 2 of Table 4.3 shows the average number of signals that experts mentioned for each scenario, while column 3 presents the average number of signals in each scenario mentioned by non-experts. A two-sample t-test between the number of signals mentioned by experts and non-experts across scenarios revealed the difference to be significant ($t = 3.3$, p-value < 0.05). To further investigate the similarities and differences, we analyzed the signals that were mentioned by one group but not the other. Column 4 shows the percentage of signals mentioned exclusively by experts (and not mentioned by non-experts), while column 5 shows signals mentioned exclusively by non-experts (and not mentioned by experts). For instance, non-experts in the "email with account" scenario mentioned fewer signals than experts. Thirty-one percent of the signals mentioned by non-experts were not considered by experts. Conversely, 38% of the signals used by experts were not mentioned by non-experts. In general, Table 4.3 shows that there are often significant differences in terms of the set of signals cited by experts and non-experts within each scenario.

We also found that most of the experts tended to mention similar signals, while non-experts tended to have a larger dispersion in the signals they mentioned. In general, we found that non-experts tended to use the structural details of an email (e.g., signature blocks such as "warm regards," security locks in the web browser, and so forth) rather than abstract principles (e.g., whether the subject is expecting an email of that type, the presence of HTTPS in the links, and so forth) to make their decision. Similar behavior was found by Chi et al. among experts and non-experts solving

Table 4.3: Average number of signals mentioned by experts and non-experts in each scenario; the values presented in the brackets are standard deviations.

Scenario	**Average number of signals mentioned by experts**	**Average number of signals mentioned by non-experts**	**Percentage of signals mentioned by experts and not mentioned by non-experts**	**Percentage of signals mentioned by non-experts and not mentioned by experts**
General email	4.5 (0.8)	2.2 (1.1)	38	31
Email with account	6.0 (1.0)	2.9 (1.5)	18	18
Email with no account	3 (0.0)	5.2 (1.6)	0	70
Email with an attachment	2.3 (1.3)	1.1 (0.9)	29	50
Auto download from website	2.6 (1.4)	2.1 (0.9)	50	37
Deliberate download from website	3.1 (0.9)	3.7 (1.2)	47	41
Buying a book online	3.1 (1.1)	2.2 (1.4)	46	31

physics problems [41]. The complete list of signals mentioned by participants for all scenarios is presented in Table 4.4.

To further understand the dispersion of signals among experts and non-experts, we calculated the correlation coefficient for each expert who mentioned a signal among all other experts who mentioned that same signal; we then calculated the mean of the coefficients across all experts. We calculated the same metrics for non-experts. We found very high correlation among experts and low correlation among non-experts across all scenarios. For instance, the mean correlation coefficient in the email with account scenario is 0.84 for experts and 0.32 for non-experts. In the email with no-account scenario, the mean correlation coefficient for experts was 1 and non-experts was 0.18. This shows that experts tend to mention the same signals, while non-experts are more dispersed in their consideration of useful signals for trust decisions. We found similar differences in the mean correlation coefficients across all other scenarios.

Table 4.4: Signals mentioned by the participants.

Scenario	**Signals mentioned by the participants**
General email	Date and time in the email; sender information; was sent to me or many people; what action to perform; size of the email; grammar or language in the email; subject information; content in the email; URL in the email; last email server; PGP signature; am I expecting the email; email header information
Email with account	Physical location of the bank; signature block in the body of the email; whether I am expecting the email; URL in the email; content of the email; email asking PII; subject information; grammatical error; last email server; sender information; header information
Email with no account	Physical location of the bank; signature block in the body of the email; whether I am expecting the email; URL in the email; content of the email; email asking PII; subject information; grammatical error; date and time of the email; sender information
Email with attachment	Date and time in the email; size of the email; size of the attachment; name of the attachment; file type of the attachment; was sent to me or many people; subject information; content of the email; sender information; what action to perform; language in the email; am I expecting the email; PGP signature; email header information
Auto download	Design of the website; trust logo on the website; time to download; size of the download; type of the download; have I heard of the download software; reputation of the website; domain name of the website; have I visited before; website using ssl
Deliberate download	Design and usability of the website; professionalism of the content in the website; URL of the website; privacy statement; time to download; size of the download; referral URL; security lock; views of friends; views of user groups; heard of the website; domain name of the website; have I visited before; reputation of the website; website using ssl; how long the website has been registered; how much do I need the software
Buying a book online	Design and usability of the website; size of the organization; privacy statement; security lock; heard of the website; price of the book; reputation of the website; website using ssl; website using https; domain name of the website; have I visited before; broken images; information from toolbars

Meaningful, misleading, or missed signals

In Section 4.2, we classified signals used in a decision-making process as meaningful, misleading, or missed.

Looking at Figure 4.3, one can evaluate how experts and non-experts differ in their usage of meaningful or misleading signals, as well as how they differ in the amount of meaningful signals they miss during their decision process. The figure compares the experts' mean *usefulness* ratings of the signals to the non-experts' mean usefulness ratings of the signals, both of which fall along a likert 1 to 7 scale. If a signal was not mentioned at all by experts or non-experts, it would appear on the 0 point with the x and y axes. The left part of the figure focuses on the email with account scenario, while the right part focuses on the email without account scenario (the patterns we detected in the figures associated with these two scenarios are similar to those associated with the other scenarios).

We observed three possible clusters of signals: those where the experts' and non-experts' opinions about the usefulness of the signals are highly correlated (signals around the 45 degree line); those where experts find signals useful but the non-experts do not, or simply ignore them (signals on the x axis or closer to the x axis); and those where non-experts find signals useful but the experts do not (signals on the y axis or closer to the y axis). Taking the experts' opinions on the utility of various signals as valid (their validity is further scrutinized below), the first category consists of what we defined in the previous sections as meaningful signals; these are depicted in Figure 4.3 as round symbols. The second category consists of what we defined as missed signals, which are depicted with square symbols; the third category consists of what we defined as misleading signals, and are depicted with triangle symbols. This representation can be considered a data-driven version of Figure 4.1 and is further analyzed in the following paragraphs.

Meaningful signals. Above, we defined "meaningful signals" as signals that enter a user's decision process and are associated with states that actually affect her well-being. Figure 4.3 focuses on two representative scenarios: "receiving an email from a bank the subject has an account with," and "receiving an email from a bank the subject has no account with." We observed that some signals mentioned by experts were also mentioned by the non-experts; these signals, shown around the 45 degree line in the Figure 4.3, could be considered meaningful signals. In particular, there is a clear difference between expert and non-expert responses in the "email with no account" scenario, where there is no signal close to the 45 degree line; this demonstrates that non-experts used only one meaningful signal (content of the email) in this scenario.

During the interviews, we found that non-experts considered a limited number of meaningful signals. For example, in the "email with no account" scenario, experts mentioned only a few signals: sender information in the email (mentioned by 100% of the experts), content of the email (mentioned by 100% of the experts) and information available in the subject line of the email (mentioned by 100%

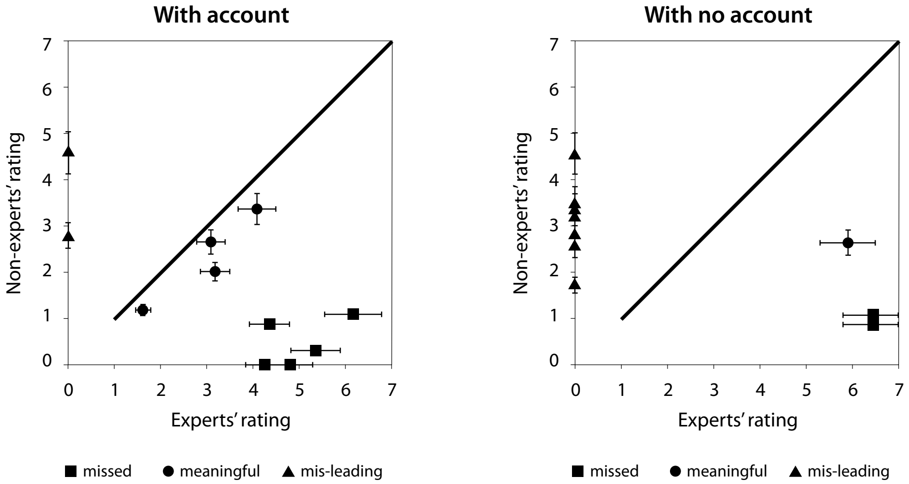


Figure 4.3: Usefulness of experts versus usefulness of non-experts. Left graph presents the email with account scenario while the right graph presents the email with no account scenario. Experts and non-experts rated the signals on a 1-7 likert scale. Circles are signals meaningful to both experts and non-experts, squares are signals missed by non-experts, and triangles are signals mis-leading to non-experts. Signals on the x axis and the y axis are completely mis-leading or missed signals; signals that are closer to the x or y axis are also considered mis-leading or missed signals. This shows that experts and non-experts disagree heavily on the usefulness of the signals.

of the experts); non-experts used the same signals only 21%, 50%, and 14% of the time.

Non-experts may not mention many meaningful signals because they lack knowledge about computer systems and security, including security indicators [53]. For instance, two meaningful signals that experts mentioned but non-experts did not mention in the "email with account" scenario were the last email server that the email came through and grammatical errors or gibberish language in the email. Instead, a typical response from a non-expert regarding the signals used while making a decision about how to deal with an incoming email was:

> "...who it is coming from. If we recognize the email address, will open it and if it is something I am expecting will open it."

A typical response from an expert was:

> "...I look for the person from whom it is coming, subject information, length or size of the email, whether there is an attachment and sometimes route and header information."

On the other hand, experts focus on more sophisticated signals when choosing their course of action. One of the experts mentioned:

> "...important emails have tags on them. They [mails] are expected to have a tag on

> the email address or it just goes out. [Email server] allows a username + and a tag @ the host and whatever comes after the plus sign is ignored by the mailer so it passes as a tag . . . [which can be used to] sort emails that way."

Misleading signals. Certain pieces of information may affect a user's decision making process even though they do not really relate to the underlying state of variables. We called tehse "misleading signals." In certain cases, the signal itself may be potentially useful, but is used in a misleading way by non-expert users. For example, the following signals have been used by non-experts to make their decisions: "professionalism of the content in the website" and "reputation and brand of the website." Non-experts are not aware that these kinds of signals can be easily spoofed. In Figure 4.3, for the "email with account" scenario, some of the signals lie in the top left of the quadrant, close to the y axis; these are signals which experts thought were not useful (or did not mention at all), but which non-experts considered highly useful. Such signals are considered "misleading signals" within the framework we have proposed.

We found that most of the signals non-experts mentioned were misleading signals in the "email with account" scenario: they were often classified as the least useful signals by the experts if not disregarded altogether. In the "email without account" scenario, all of the signals that experts rated low are rated high by the non-experts. For instance, non-experts mentioned the physical location of the bank that sent the email and the email signature block (e.g. "Warm regards"), while experts did not. These findings are duplicated for most of the signals mentioned by non-experts and experts across other scenarios: in the "buying" scenario, non-experts mentioned the security lock on the website and the usability of the website, while experts did not. In the "auto download from website" scenario, one misleading signal was the time it takes to download the software.

Missed signals. In this model, useful data ignored by users are defined as "missed signals." During the interviews, we observed that none of the following signals in the email scenarios were mentioned by non-experts: length or size of the email, last email server, and PGP signature. However, most of these signals were frequently mentioned by experts. In addition, the following signals were not mentioned by non-experts in the website scenarios: HTTPS, broken images, information presented in the status bar, Secure Sockets Layer (SSL), "whois" information for the website, and phishing toolbars. These missed signals appear as squares along the horizontal axis for both scenarios in Figure 4.3. Non-experts typically made decisions in the "email with account" scenario based solely on the sender's email address, and were often unaware of many other signals which could have been more useful than the ones they used. We also observed that they were not aware that most of the signals they mentioned can be spoofed.

In fact, across various scenarios, we often found that most signals experts considered highly useful were considered not very useful by non-experts, and vice versa: the signals that experts rate low are rated high by the non-experts. Results therefore suggest that there is a large difference between experts and non-experts in the average number and types of signals used to make decisions. In general, we found that non-experts tended to focus more on the structural details of an email rather than on abstract principles when making decisions. These results also suggest that the number of meaningful signals that non-experts consider is limited, that they use far more misleading signals than experts, and that they miss some of the signals frequently mentioned by experts.

Above, we have implicitly assumed that the experts' opinions on the utility of various signals is the correct view, and have measured them against the non-experts' opinions. To validate the *actual* usefulness of different signals and vet this reliance on experts' opinions, we asked a separate group of five security researchers conducting phishing research at Carnegie Mellon University to rank the signals that experts and non-experts had mentioned during the interviews according to their usefulness in helping users make more accurate decisions. Usefulness was measured on a likert scale from 1 to 7 (where 1 represents "not at all useful" and 7 represents "most useful"). This analysis was performed after we had finished collecting the data from the experts and non-experts. We will refer to these researchers as "super experts" to differentiate them from the other groups. Super experts tended to largely agree with each other on the ratings of the usefulness of the signals. The mean correlation coefficient for the usefulness of the signals among super experts was 0.72 (the coefficient was calculated using the usefulness likert value that super experts mentioned). Using the data we collected from super experts, we performed a linear regression analysis, using (separately) experts' and non-experts' usefulness ratings as the dependent variable in regressions over the super experts' ratings as the independent variable. Table 4.5 shows that if super experts gave a high usefulness rating to a signal, non-experts gave it a low rating (negative and significant coefficient), while super experts and experts agreed on the usefulness of the signals (positive and significant coefficient).

Table 4.5: Regression analysis for the signal usefulness ratings given by experts and non-experts compared to those given by super experts. The results show that experts and super experts agree on ratings, while non-experts disagree. ** indicates significance at the 1% level; * indicates significance at the 5% level.

	With account		Without account	
	Expert	Non-expert	Expert	Non-expert
Intercept	2.67^{**} (0.96)	5.4^{**} (3.9)	1.78^{**} (0.15)	8.3^{*} (2.95)
Signal usefulness	1.1	-0.65	1.1	-0.24
R^2	0.95	0.62	0.99	0.37

4.4.3 Actions

Actions are the set of activities that a user may perform in a certain scenario. To collect data regarding actions, we asked the question: "What are all the different things that you could do in this situation?" Klein et al. show that experts usually select the best option as their first option when making a decision, and thereby avoid the need to perform extensive generation and evaluation [105]. We found similar results in the data. For example, in the "email with no account" scenario, all of the experts mentioned deleting the email as the first action, while only 15% of the non-experts mentioned deleting the email as the first option.[1] Similarly, in the "email with attachment" scenario, we found that 73% of the experts mentioned deleting the email as one of their first actions versus 29% among non-experts (Table 4.6). The complete list of all actions mentioned by participants is presented in Table 4.7.

Column 2 of Table 4.6 presents the average number of actions mentioned by experts, while column 3 presents the average number of actions mentioned by non-experts. Columns 4 and 5 present the percentage of actions mentioned by experts and not by non-experts, and actions mentioned by non-experts and not by experts, respectively. The average and percentage discussed here were calculated in the same manner as in Section 4.4.2, but using the data from questions related to actions. Performing a two sample t-test on the average number of signals mentioned by experts and non-experts, we found no statistically significant difference (t = 0.65, p-value = 0.52). However, within each scenario, the mean correlation coefficient across the actions mentioned by experts tended to be high, whereas the correlation coefficient across the actions mentioned by non-experts tended to be very low (for instance, for the "email with account" scenario, the correlation coefficient among experts was 0.9, but only 0.12 among non-experts). This result confirms that experts tend to agree on the set of actions to consider in trust sensitive situations, but non-experts' set of actions are more dispersed. Furthermore, we again found a significant number of actions mentioned by experts but not by non-experts and vice versa, indicating that non-experts may be engaging in actions that are not necessarily appropriate. More broadly, while higher skill levels are very useful in selecting the first action (for example, in the "email with no account" scenario, experts chose "delete email" as the first option), there is no strong relationship between skill level and the rest of the actions that experts and non-experts take. These results are in line with findings in the literature on experts (see, for instance, [105]).

[1] By 'first' action we refer to the action that a participant mentioned first, after the question regarding the action was asked. As noted above, there is no guarantee that what participants claimed they would do in a given scenario would match what they would actually do in the real scenario. Analysis is based on a mental model methodology, focusing on experts and non-experts' awareness and reliance of the existence and value of information in online trust scenarios. In related studies [109], we focused on the analysis of behavior in controlled experiments.

Table 4.6: Average number of actions mentioned in each scenario; values presented in the brackets are standard deviations. There is no significant difference between experts and non-experts in the average number of actions they mentioned.

scenario	Average number of actions mentioned by experts	Average number of actions mentioned by non-experts	Percentage of actions mentioned by experts and not mentioned by non-experts	Percentage of actions mentioned by non-experts and not mentioned by experts
General email	2.2 (0.8)	2.2 (0.8)	9	36
Email with account	1.8 (0.8)	1.6 (0.6)	22	33
Email with no account	1.4 (0.8)	1.7 (0.7)	14	57
Email with an attachment	2 (0.4)	1.7 (0.7)	13	0
Auto download from website	2 (0.4)	1.6 (0.9)	33	17
Deliberate download from website	1.6 (0.5)	1.6 (0.7)	50	17
Buying a book online	1.9 (0.7)	1.7 (0.6)	17	17

4.4.4 Attackers

In order to understand the knowledge of experts and non-experts about the possible influence of third parties on what signals are visible to them, we asked "Is there any third person who might change these things that you look for while making your decision? If so, who are they?" A common response from the non-experts was "I don't think so." Only around 50% of the non-experts thought that a third party could influence the signals they used in their decision process in any of the seven scenarios (across all scenarios, non-experts only mentioned attackers about half of the time). All experts in the study mentioned that criminals could intercept and manipulate anything. One common expert response was, "Spammers, phishers, virus writers."

Table 4.7: Actions mentioned by the participants.

Scenario	**Actions mentioned by the participants**
General email	Open the email; read the email; reply to the email; delete the email; ignore the email; mark it as spam; download the attachment; open the email in preview pane; print the email; move it to a folder; do the action mentioned in the email
Email with account	Read the email; open the email; do not click on the link in the email; delete the email; call or go to the bank; ignore the email; print the email; click on the link in the email
Email with no account	Open the email; ignore the email; read the email; reply to the email; delete the email; mark it as spam; click on the link in the email
Email with attachment	Download the attachment; do not download the attachment; delete the email; open the email; reply to the email; read the email; ignore the email; mark the email as spam; download the attachment and run virus scanner
Auto download	Download the software; will not download the software; use a different channel to communicate with the website; close the window; will search to find more information about the download; come back to website later; check from another website
Deliberate download	Download the software from reputed website; will not download it from a non-reputed site; open a new browser and type the URL to download; ask friends to make a decision from where to download; download and run the virus scanner; search to find the appropriate download
Buying a book online	Give personal information to buy the book; find a reputable website to buy the book; will not buy it from a non-reputable website; will not buy the book; search to find appropriate website; go to a shop and buy the book; will do comparison shopping and then make a decision

4.5 Discussion

Results suggest that, on average, non-experts use fewer meaningful and more misleading signals than experts, and that non-experts also miss many useful signals. Findings also indicate that non-experts are often unaware of how criminals can impersonate legitimate organizations and get personal information from victims.

These results emphasize the value of and need for user education, and may be used to inform people designing methods that will be used to educate non-experts. Future educational efforts may want to focus on providing information about how criminals operate online, making non-experts more aware of the problems associated with some of the signals they use, and helping non-experts understand that other signals are available. In particular, one way to improve non-experts' decision processes may be to increase their awareness of attacks, means of protection, and signals revealing something about the true nature of an email or website. Non-experts should be provided with both declarative knowledge (e.g. what a phishing attack is) and procedural knowledge (e.g. how to identify phishing emails) [10]. They should also be taught basic Internet safety rules like "don't believe everything you read," "a polished appearance is not the same as substance," and "if something is too good to be true, it probably is" [116]. They should also be provided with simple but specific instructions such as "type the real website address into a web browser," and "never give out personal information upon email request" [109]. Similar suggestions for educating users have been provided by [18,35,55,117].

Clearly, making non-experts into experts is an unattainable goal, but offering non-experts practical knowledge to help them make more informed decisions is quite feasible. But educators must do more than provide reading materials; in one of the laboratory studies (see Section 5.3), we found that people don't read security notices (emails that organizations send out with a link to training materials) sent to them. Results from this study informed the design and development of PhishGuru. We describe the rationale and design evolutions in the next chapter and also show the effectiveness of PhishGuru in Chapter 6 and Chapter 7.

Chapter 5

PhishGuru Methodology and Interventions

In this Chapter we discuss the design of the embedded training system called PhishGuru.[1] In Section 5.1, we discuss the results of a study in which we evaluated existing online training materials. In Section 5.2, we describe the design rationale for the embedded training concept. In Section 5.3, we discuss the evolution of the PhishGuru design and discuss the iterative evaluations we performed.

5.1 Evaluation of existing online training materials

> This section is joint work with Steve Sheng, Alessandro Acquisti, Lorrie Cranor, and Jason Hong, and published as a CyLab technical report [112].

In Chapter 3, we discussed user studies from the literature that measured the effectiveness of specific training materials. However, those studies did not analyze the quality of the training materials being tested or consider ways of designing more effective training materials. In this section, we present an analysis using the instructional design principles described in Section 2.3 on online training materials; we also present the results of a user study examining the effectiveness of existing training materials.

5.1.1 Training material selection and analysis

To analyze the online training materials, we selected three representative tutorials from well-known sources: eBay's tutorial on spoofed emails [56], Microsoft's security tutorial on Phishing [137], and

[1] http://www.phishguru.org/

the Phishing E-card from the U.S. Federal Trade Commission [62]. Since none of these tutorials provide much information on parsing URLs—a skill that can help people identify fraudulent links—we also selected a URL tutorial from the online security education portal MySecureCyberspace [149].

Table 5.1 presents information about the format and length of the training materials we evaluated, summarizing the concepts taught by each. Most of the training materials we examined present a basic definition of phishing, highlight common characteristics of phishing emails, provide suggestions to avoid falling for these scams, and offer information about what to do after falling for them. The materials also provide a link to other resources about phishing and security. A common message in most of these materials was that trusted organizations will not request personal information through email.

Table 5.1: Characteristics of selected online training materials.

Source	Format	Length			Concepts taught	
		Words	**Printed pages**	**Graphic examples**	**Cues to look for**	**Guidelines**
Microsoft	Web page	737	3	2	Urging urgent action; non-personalized greeting; requesting personal information	Identify fraudulent links
eBay	Web page	1276	5	8	all the above; sender email address; links in the email; legitimate vs. fake eBay address	Never click on links in email; identify fraudulent links
FTC	Video	N/A	N/A	N/A	Requesting personal information	Never respond to emailed requests for personal information
MySecure Cyberspace	Web page	236	1	0	N/A	N/A

The training materials we selected made minimal use of the basic instructional design principles introduced in Table 2.1. The eBay and Microsoft tutorials used the contiguity principle, while the FTC video used the personalization and story-based agent environment principles. We found that illustrations were used more for decorative purposes than explanative purposes, and that those used for explanative purposes sometimes lacked captions or explanations in the body of the text. In some cases text and associated images were located far from each other, either much farther down on a long web page or on a different web page altogether.

5.1.2 User study

We conducted a user study to evaluate the effectiveness of the online training materials. Fourteen participants were asked to examine 10 websites and determine which were phishing sites. They were then given 15 minutes to read the four selected training materials. After training, the participants were asked to examine 10 more websites and determine which were phishing sites. A control group of fourteen participants completed the same protocol, but spent the 15-minute break playing solitaire and checking their email instead of reading training materials.

We measured false positives and false negatives before and after training. A false positive occurs when a legitimate website is mistakenly judged to be a phishing website. A false negative occurs when a phishing website is incorrectly judged to be legitimate. We found that false negatives fell from 38% before training to 12% after training. However, false positives increased from 3% to 41%. The control group did not perform significantly differently before and after their 15-minute break. Further details of this study are discussed in Section 8.2.2.

The results suggest that the existing online training materials are surprisingly effective at helping users identify phishing websites when users actually read the training materials. However, they could also be made more effective by applying basic instructional design principles. Furthermore, while the results demonstrate that users are better at avoiding phishing websites after reading the training materials, users are also more likely to have false positives. Finally, even when more effective online training materials are available, getting users to read them voluntarily remains a challenge.

The rest of this chapter describes the work we have done to develop better approaches to anti-phishing training through principled instructional design, innovative delivery methods, and effective content for the training materials.

5.2 Design of the PhishGuru concept

Education researchers believe that training materials are most effective when they incorporate the context or situation of the real-world, work, or testing situation [13,44]. With embedded training, training materials are integrated into the primary tasks users perform in their day to day lives. This training method has been widely applied in the training of military personnel on new Future Combat Systems (FCS) [30,103].

There are two primary intervention points for an anti-phishing training system: the email and the website. We chose to focus on emails rather than websites for three reasons. First, email is the main vector for delivering phishing messages to users. If we can prevent people from trusting phishing emails, they will not visit phishing websites. Second, anti-phishing websites [56,62] require end-users to proactively visit them, limiting the number of people who actually see them. In contrast, an embedded training approach brings information to end users and teaches them over time to differentiate between legitimate and illegitimate emails. Third, end users must already have some knowledge about phishing or other kinds of scams to seek out educational websites. In contrast, embedded training (if distributed with standard email clients or sent by companies) works for experts as well as non-experts who are unaware of phishing; it does this by educating end-users immediately after they have made a mistake. Studies have shown that providing immediate feedback enhances learning [10, Chapter 7], [126].

In the PhishGuru approach, people are periodically sent training emails, perhaps from their system administrator or a training company. People access these training emails in their inbox when they check their regular emails. The training emails look just like phishing emails, urging people to go to some website and log in. If people fall for the training email – that is, if they click on a link in that email – we provide an intervention message that explains that they are at risk for phishing attacks and offers tips users can follow to protect themselves. Providing immediate feedback at this "teachable moment" enhances learning [10], [126]. There is a plethora of literature on *teachable moments* in fields such as sexual behavior and HIV prevention, injury prevention, and smoking cessation [131]. When users click on a link in a PhishGuru training email, they encounter a training message that alerts them to the risk of clicking on links, thereby creating a teachable moment that can influence user behavior. This approach has the following advantages: (1) it enables a system administrator or training company to continually train people as new phishing methods arise; (2) it enables users to be trained without taking time out of their busy schedules (since the training is part of a primary task); and (3) it creates a stronger motivation for users, as training materials are presented only after they actually "fall" for a phishing email.

One early design consideration was whether to show interventions immediately after a person clicked on a training email or after they tried to log into the phishing website. A pilot test with paper prototypes strongly suggested that showing an intervention after a person had clicked on a link was better, since people who were shown interventions after logging in were confused as to why they were seeing warning messages about the risks of clicking on email links. We believe this is due to a gap between cause (clicking on a link) and effect (seeing a warning message about email after logging in). Egelman et al. observed a similar gap when user study participants who had seen a web browser anti-phishing warning returned to the email that triggered the warning and repeatedly tried to access the fraudulent website, unaware that the email itself was fraudulent [58]. In a laboratory study (discussed in Section 6.2), we found that users retained more knowledge when they received embedded training than when the intervention was sent as an email (non-embedded). Therefore, we decided to embed the interventions, presenting them at the moment users click on the link within emails.

We applied the instructional design principles discussed in Section 2.3.1 to the design of the PhishGuru. We applied the conceptual-procedural principle by defining phishing (conceptual knowledge) at the top of the design (Figure 5.1), and presenting ways to protect oneself (procedural knowledge) at the right-hand side. Table 5.2 summarizes the ways in which we applied instructional design principles when designing the PhishGuru.

Table 5.2: Application of instructional design principles to the design of PhishGuru.

Principle	**Way(s) in which we applied the principle in our design**
Learning-by-doing	In our approach, users learn by actually clicking on phishing emails (doing). The training materials are presented when users fall for phishing emails.
Immediate feedback	We provide feedback through interventions immediately after the user clicks on a link in a fake phishing email sent by us.
Conceptual-procedural	In the top strip of the design (Figure 5.1) we define phishing (conceptual knowledge). On the right-hand side, we present ways to protect (procedural knowledge) oneself from phishing.
Contiguity	We have placed pictures and relevant text contiguously in the instructions (numbered 1 through 6 in Figure 5.1).
Personalization	We apply this principle in the design in many ways. For example, characters say things like “I forged the address to look genuine” (Figure 5.1).
Story-based agent environment	We have three characters: the phisher, the victim, and the PhishGuru. They are placed in Figure 5.1 with the story in the background.

5.3 Evolution of PhishGuru interventions

To gain insight into the design space, we created and evaluated several prototypes of the interventions. We started with paper prototypes and refined the ideas using HTML prototypes. To get a better feel for how the PhishGuru idea would work in practice, we created an HTML mockup in SquirrelMail, a web-based email service. People who used the system encountered training emails interspersed with regular email messages. If they clicked on a link in one of the training emails, they were taken to a separate web page and shown one of two interventions. The first intervention (see Figure 5.2) showed a screenshot of the email within the web browser itself, pointing out that the link the user clicked on was not the same as the link they would actually go to as shown in the

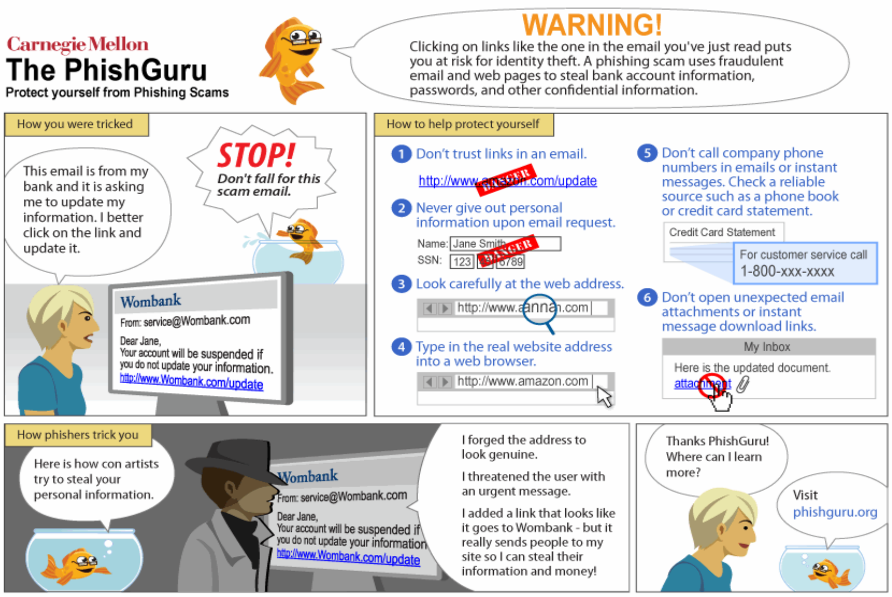

Carnegie Mellon
The PhishGuru
Protect yourself from Phishing Scams
WARNING!
Clicking on links like the one in the email you've just read puts you at risk for identity theft. A phishing scam uses fraudulent email and web pages to steal bank account information, passwords, and other confidential information.
How you were tricked
This email is from my bank and it is asking me to update my information. I better click on the link and update it.
STOP! Don't fall for this scam email.
Wombank
From: service@Wombank.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.Wombank.com/update
How to help protect yourself
1 Don't trust links in an email.
http://www.amazon.com/update
DANGER
2 Never give out personal information upon email request.
Name: Jane Smith
SSN: 123 45 6789
DANGER
3 Look carefully at the web address.
http://www.amazon.com
4 Type in the real website address into a web browser.
http://www.amazon.com
5 Don't call company phone numbers in emails or instant messages. Check a reliable source such as a phone book or credit card statement.
Credit Card Statement
For customer service call 1-800-xxx-xxxx
6 Don't open unexpected email attachments or instant message download links.
My Inbox
Here is the updated document.
attachment
How phishers trick you
Here is how con artists try to steal your personal information.
Wombank
From: service@Wombank.com
Dear Jane,
Your account will be suspended if you do not update your information
http://www.Wombank.com/update
I forged the address to look genuine.
I threatened the user with an urgent message.
I added a link that looks like it goes to Wombank - but it really sends people to my site so I can steal their information and money!
Thanks PhishGuru! Where can I learn more?
Visit phishguru.org

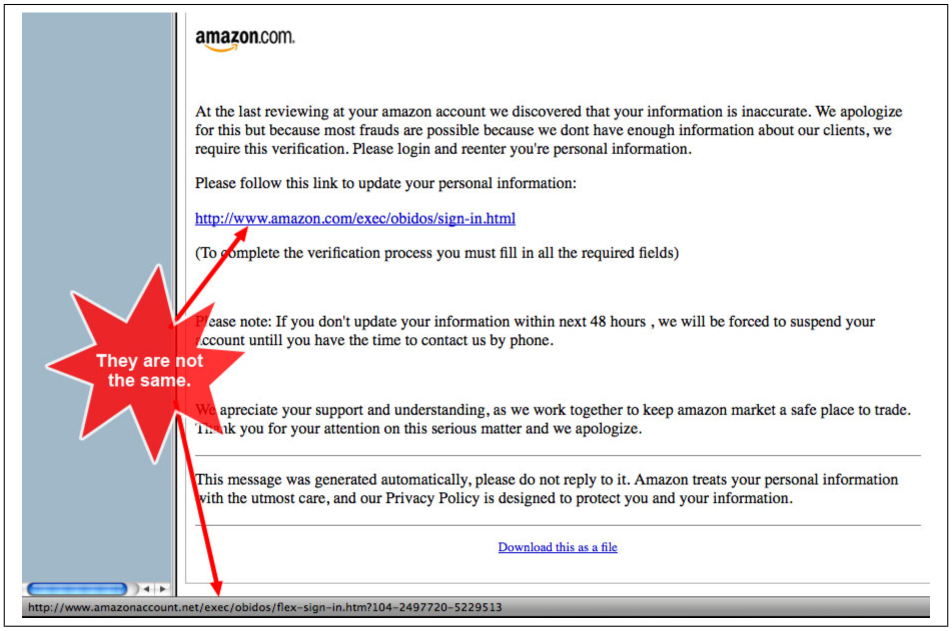
Figure 5.1: This is the final intervention design that we used in a large scale real world study [108].

status bar. This intervention provides only one instruction, pointing out that the URL in the email and the target URL in the status bar are not the same. The second intervention (see Figure 5.3) was similar, but told people more directly that the link they clicked on did not take them to the intended website; it did this by displaying the brand name itself (in this case, “This is not eBay”). Both interventions also provided text at the top of the image describing why the participants were seeing such a page and informing them that they were susceptible to phishing attacks.

We did a pilot evaluation of the intervention with ten participants, using a variation of the protocol developed by Downs et al [54]. We asked participants to role play as an employee at a company and to handle the email in the employee’s mailbox the way they normally would. The employee’s mailbox contained nineteen email messages; among the nineteen were a few phishing emails and two training emails.

Nine out of ten participants clicked on the first training message (falling for the fake phishing email) and saw the information we presented about phishing. However, almost of all the users who viewed the training message were confused about what was happening. They did not understand why they had been sent this email.

Furthermore, most of the participants who viewed the training message did not understand what it was trying to convey. A common response to the first intervention (Figure 5.2) was, “I don’t know what it is trying to tell me.” Some users understood the training message but were uncertain how to respond, as the message did not suggest any specific actions to take. In debriefing sessions, participants reported that the second intervention was more useful than the first, since they could



At the last reviewing at your amazon account we discovered that your information is inaccurate. We apologize for this but because most frauds are possible because we dont have enough information about our clients, we require this verification. Please login and reenter you're personal information.

Please follow this link to update your personal information:

http://www.amazon.com/exec/obidos/sign-in.html

(To complete the verification process you must fill in all the required fields)

Please note: If you don't update your information within next 48 hours , we will be forced to suspend your account untill you have the time to contact us by phone.

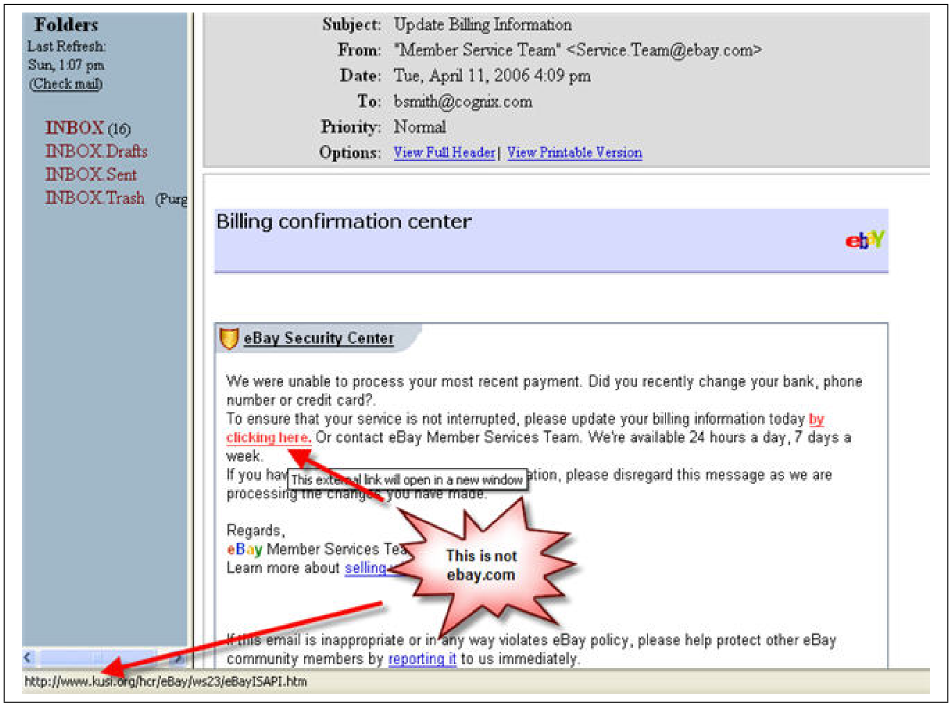
We apreciate your support and understanding, as we work together to keep amazon market a safe place to trade. Thank you for your attention on this serious matter and we apologize.

This message was generated automatically, please do not reply to it. Amazon treats your personal information with the utmost care, and our Privacy Policy is designed to protect you and your information.

Download this as a file

http://www.amazonaccount.net/exec/obidos/flex-sign-in.htm?104-2497720-5229513

Figure 5.2: First intervention with only one instruction – comparing the URL in the email and the URL in the status bar.



Subject: Update Billing Information
From: "Member Service Team" <Service.Team@ebay.com>
Date: Tue, April 11, 2006 4:09 pm
To: bsmith@cognix.com
Priority: Normal
Options: View Full Header | View Printable Version

Billing confirmation center

eBay Security Center

We were unable to process your most recent payment. Did you recently change your bank, phone number or credit card?.
To ensure that your service is not interrupted, please update your billing information today by clicking here. Or contact eBay Member Services Team. We're available 24 hours a day, 7 days a week.
If you hav ation, please disregard this message as we are processing the changes you have made.

Regards,
eBay Member Services Tea
Learn more about selling

If this email is inappropriate or in any way violates eBay policy, please help protect other eBay community members by reporting it to us immediately.

http://www.kusi.org/hcr/eBay/ws23/eBayISAPI.htm

Figure 5.3: eBay intervention with only one instruction – emphasizing that the URL in the email and the target URL of the link in the email are not same.

understand that the website they were visiting was not part of eBay.

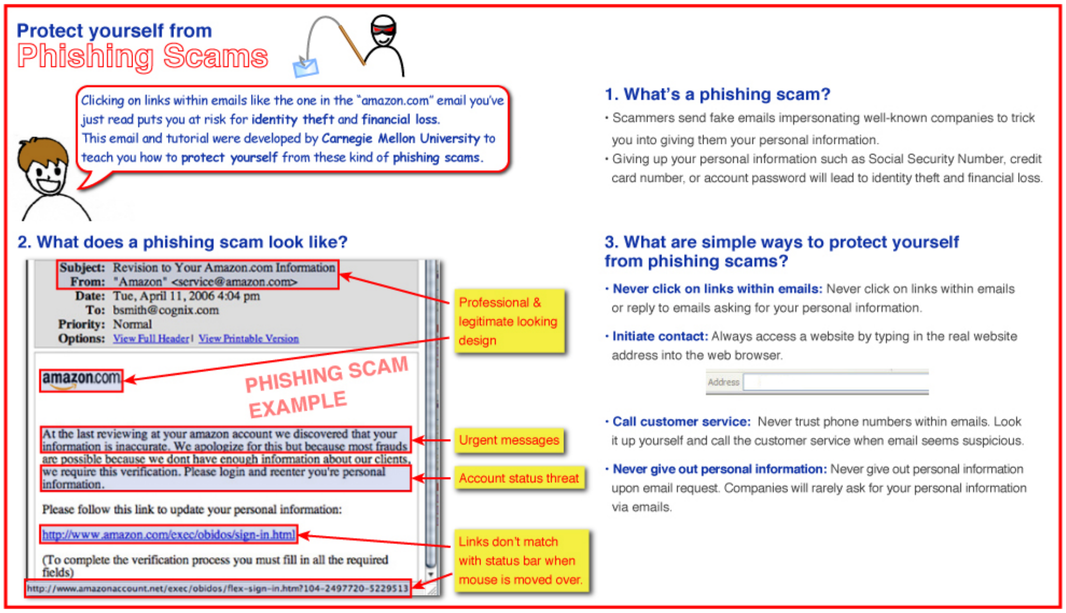
Another flaw of the design was that people were sometimes confused by the screenshot of the web browser. Many participants failed to notice the text at the top describing why they were seeing the warning, mostly because the browser screenshot was so large and visually dominant. A third drawback was that people had to scroll to see the entire warning.

Nine users fell for the first phishing email (before any interventions), and seven users fell for the final phishing email (after both interventions), suggesting that this early design was not effective. Nearly all of the participants who clicked on a phishing link actually tried to log in. This again suggests that it would be better to intervene immediately after a person clicks on a link (since they are likely to fall for the phishing website) rather than after they try to log in. In summary, the lessons from early prototypes were:

1. It is best to show interventions immediately after a person clicks on a link in a training email.
2. Since people expect to go to a website when they click on a link, interventions need to make it extremely clear why they are not being taken to that website.
3. Interventions need to provide clear actionable items rather than general warnings about potential risks.
4. Text and images need to be simple and visually salient to convey the warning accurately and avoid confusion.
5. Text segments need to be brief.

To implement what we learned and further develop the content, we compiled a list of 25 online anti-phishing training materials. In addition, we consulted with experts to select a set of frequently mentioned guidelines, guidelines that offered simple and effective steps end users could take to avoid falling for phishing attacks. We eliminated guidelines that focused on strategies that would be difficult for many users, such as using networking tools to determine the age and owner of a domain. Based on this analysis, we selected the following four guidelines: (1) Never click on links in emails; (2) Initiate contact (i.e. manually type URLs into the web browser); (3) Call customer service; (4) Never give out personal information.

The first guideline was somewhat controversial among the experts we consulted. While they agreed that users who do not click on links will not be susceptible to most email-based phishing attacks, some experts argued that email links offer considerable convenience and value to users. As such, they argued that it would be unrealistic for users to stop clicking on all links in emails. Therefore, it is important to teach users how to identify and avoid clicking on links likely to lead to fraudulent websites. However, the process of identifying fraudulent links is complex. The rationale for the



Protect yourself from
Phishing Scams

Clicking on links within emails like the one in the "amazon.com" email you've just read puts you at risk for **identity theft** and **financial loss**.
This email and tutorial were developed by **Carnegie Mellon University** to teach you how to **protect yourself** from these kind of **phishing scams**.

1. What's a phishing scam?

- Scammers send fake emails impersonating well-known companies to trick you into giving them your personal information.
- Giving up your personal information such as Social Security Number, credit card number, or account password will lead to identity theft and financial loss.

2. What does a phishing scam look like?

Subject: Revision to Your Amazon.com Information
From: "Amazon" <service@amazon.com>
Date: Tue, April 11, 2006 4:04 pm
To: bsmith@cognix.com
Priority: Normal
Options: View Full Header | View Printable Version

amazon.com

At the last reviewing at your amazon account we discovered that your information is inaccurate. We apologize for this but because most frauds are possible because we dont have enough information about our clients, we require this verification. Please login and reenter you're personal information.

Please follow this link to update your personal information:

http://www.amazon.com/exec/obidos/sign-in.html

(To complete the verification process you must fill in all the required fields)

http://www.amazonaccount.net/exec/obidos/flex-sign-in.htm?104-2497720-5229513

PHISHING SCAM EXAMPLE
Professional & legitimate looking design
Urgent messages
Account status threat
Links don't match with status bar when mouse is moved over.

3. What are simple ways to protect yourself from phishing scams?

- **Never click on links within emails:** Never click on links within emails or reply to emails asking for your personal information.
- **Initiate contact:** Always access a website by typing in the real website address into the web browser.
- **Call customer service:** Never trust phone numbers within emails. Look it up yourself and call the customer service when email seems suspicious.
- **Never give out personal information:** Never give out personal information upon email request. Companies will rarely ask for your personal information via emails.

Figure 5.4: First text/graphic intervention. This defines phishing, and lists simple ways to protect yourself from phishing scams. It also provides an example of the phishing email annotated with features in the email. Used in a laboratory study [109].

second guideline, "Initiate contact," is that it is much safer for people to type a web address into a web browser on their own than for them to trust a link in an email. For the third guideline, "Call customer service," the rationale is that many phishing attacks rely on scaring people into logging in to an account. Calling customer service is a fairly reliable way to determine if there really are any problems with one's account (assuming the phone number is obtained from a reliable source). We also believe that a higher number of customer service calls will encourage companies to take stronger action against phishing, since such calls cost companies money. Although this seems like an extreme measure, it is worth noting that no person in the studies actually called customer service. We argue that this is still a useful piece of advice, as it reminds people that there are offline ways to contact companies. For the fourth guideline, "Never give out personal information," the rationale is that companies rarely ask for such information, and that the large majority of such requests are phishing attacks.

Informed by early designs, we created two new interventions: a text/graphic intervention and a comic strip intervention. The text/graphic intervention, shown in Figure 5.4, describes the risks of phishing, shows a small screenshot of the training email, points out cues that identify it as a phishing email, and outlines simple actions that users can take to protect themselves. The comic strip intervention, shown in Figure 5.5, conveys roughly the same information as the text and graphics intervention, but in a comic strip format. The rationale here was that the first intervention had a great deal of text, which might cause people to close the window without reading it. Since

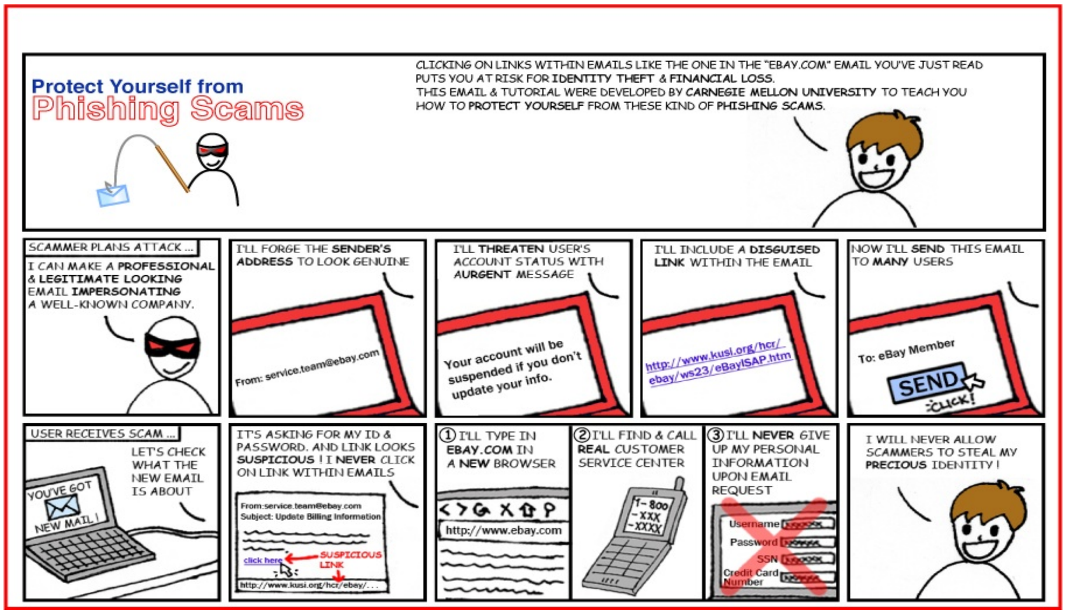

Protect Yourself from
Phishing Scams
CLICKING ON LINKS WITHIN EMAILS LIKE THE ONE IN THE "EBAY.COM" EMAIL YOU'VE JUST READ PUTS YOU AT RISK FOR IDENTITY THEFT & FINANCIAL LOSS.
THIS EMAIL & TUTORIAL WERE DEVELOPED BY CARNEGIE MELLON UNIVERSITY TO TEACH YOU HOW TO PROTECT YOURSELF FROM THESE KIND OF PHISHING SCAMS.
SCAMMER PLANS ATTACK ...
I CAN MAKE A PROFESSIONAL & LEGITIMATE LOOKING EMAIL IMPERSONATING A WELL-KNOWN COMPANY.
I'LL FORGE THE SENDER'S ADDRESS TO LOOK GENUINE
From: service.team@ebay.com
I'LL THREATEN USER'S ACCOUNT STATUS WITH AURGENT MESSAGE
Your account will be suspended if you don't update your info.
I'LL INCLUDE A DISGUISED LINK WITHIN THE EMAIL
http://www.kusi.org/hcr/ebay/ws23/eBayISAP.htm
NOW I'LL SEND THIS EMAIL TO MANY USERS
To: eBay Member
SEND
CLICK!
USER RECEIVES SCAM ...
LET'S CHECK WHAT THE NEW EMAIL IS ABOUT
YOU'VE GOT NEW MAIL!
IT'S ASKING FOR MY ID & PASSWORD. AND LINK LOOKS SUSPICIOUS ! I NEVER CLICK ON LINK WITHIN EMAILS
From:service.team@ebay.com
Subject: Update Billing Information
click here
SUSPICIOUS LINK
http://www.kusi.org/hcr/ebay/...
① I'LL TYPE IN EBAY.COM IN A NEW BROWSER
http://www.ebay.com
② I'LL FIND & CALL REAL CUSTOMER SERVICE CENTER
1-800-XXX-XXXX
③ I'LL NEVER GIVE UP MY PERSONAL INFORMATION UPON EMAIL REQUEST
Username
Password
SSN
Credit Card Number
I WILL NEVER ALLOW SCAMMERS TO STEAL MY PRECIOUS IDENTITY !

Figure 5.5: This is the first comic strip intervention. This presents the same information as Figure 5.4, but in a comic strip format. [109].

comic strips stories are a highly approachable medium [44], so we decided to test the effectiveness of a comic strip approach to anti-phishing training. We evaluated these two designs through a laboratory study; the results showed that users who saw the comic strip version did better than people who saw the text/graphic version. This helped us narrow the design space to comic strip interventions. Both interventions use prominent titles and a cartoon image of a thief to convey that participants are potentially at risk. We designed the interventions so they could be read without scrolling or clicking on additional links. Keeping the one page constraint in mind, all PhishGuru interventions are designed to fit on one page.

Using feedback from the laboratory study and our understanding of the literature, we tried to create a stronger agent and a story we could use in the interventions. After a few iterations (getting users' feedback), we arrived at a character for the intervention (see for the man with turban in Figure 5.6) to be called "PhishGuru." To emphasize the story based principle, we created the victim character, who is instructed by the PhishGuru to avoid falling for phishing attacks. The intervention also includes the phisher character, who sends phishing emails to victims. We also applied the contiguity principle by providing an illustration to go with each instruction. We added the instruction "Always be wary of suspicious websites." The rationale for this instruction was to inform users that they should be cautious about domain names (for example similar looking names like *annazon* instead of *amazon*). We used this intervention in a study where we evaluated retention and transfer of knowledge through PhishGuru training methodology [110].

We received some criticism for the PhishGuru character representing a specific ethnic group; there-

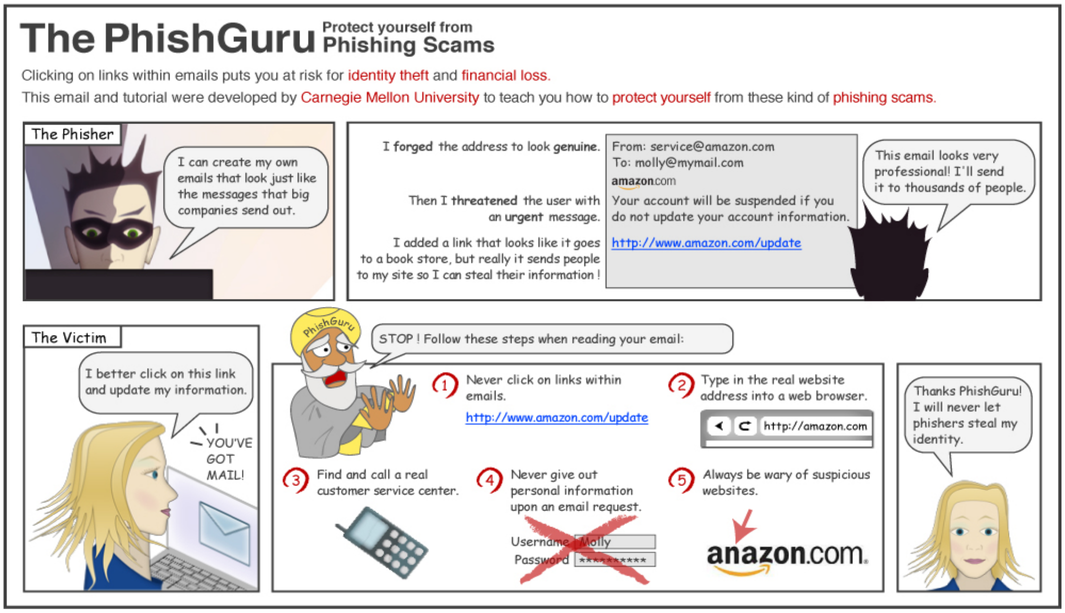

The PhishGuru Protect yourself from Phishing Scams
Clicking on links within emails puts you at risk for identity theft and financial loss.
This email and tutorial were developed by Carnegie Mellon University to teach you how to protect yourself from these kind of phishing scams.
The Phisher
I can create my own emails that look just like the messages that big companies send out.
I forged the address to look genuine.
Then I threatened the user with an urgent message.
I added a link that looks like it goes to a book store, but really it sends people to my site so I can steal their information !
From: service@amazon.com
To: molly@mymail.com
amazon.com
Your account will be suspended if you do not update your account information.
http://www.amazon.com/update
This email looks very professional! I'll send it to thousands of people.
The Victim
I better click on this link and update my information.
YOU'VE GOT MAIL!
PhishGuru
STOP ! Follow these steps when reading your email:
1 Never click on links within emails.
http://www.amazon.com/update
2 Type in the real website address into a web browser.
http://amazon.com
3 Find and call a real customer service center.
4 Never give out personal information upon an email request.
Username Molly
Password **********
5 Always be wary of suspicious websites.
amazon.com
Thanks PhishGuru! I will never let phishers steal my identity.

Figure 5.6: This is an updated version of Figure 5.5. In this version we applied the contiguity principle to all instructions. We also added information about what phishers do to send phishing emails. Used in a laboratory study [110].

fore, we created a gender- and ethnic- neutral character – a fish (see Figure 5.7). We used this intervention in a real world study conducted among employees of a large ISP in Portugal [113].

Using feedback from the study and other input, we modified the intervention to Figure 5.8. We conducted two focus group studies to evaluate the content of PhishGuru (further details of this focus group are discussed in Section 8.1.3.) In the focus groups, participants of all ages mentioned that they would read the entire PhishGuru intervention over other training materials. Participants also did not like the all-capital comic font in the intervention. Most of the participants in the focus group did not like the phisher character either.

Using feedback from the focus groups, we modified the intervention to Figure 5.1. Mainly, we changed the font, created a new phisher character, and cleaned up the language. To teach users to avoid opening attachments or clicking on links within instant messages, we added the instruction "Don't open unexpected email attachments or instant message download links." Since some participants of the focus group noted that villains in the interventions (e.g. the phisher) are almost always male, we created a version of the comic strip in which the phisher was female (See Figure 5.9). In this intervention, we also varied the story so that users who view the intervention multiple times do not have to keep reading the same one or get habituated and disregard the intervention. This intervention has exactly the same instructions as Figure 5.1. We used these two interventions in a real world study conducted among Carnegie Mellon's students, staff, and faculty. We discuss this study in detail in Section 7.2.

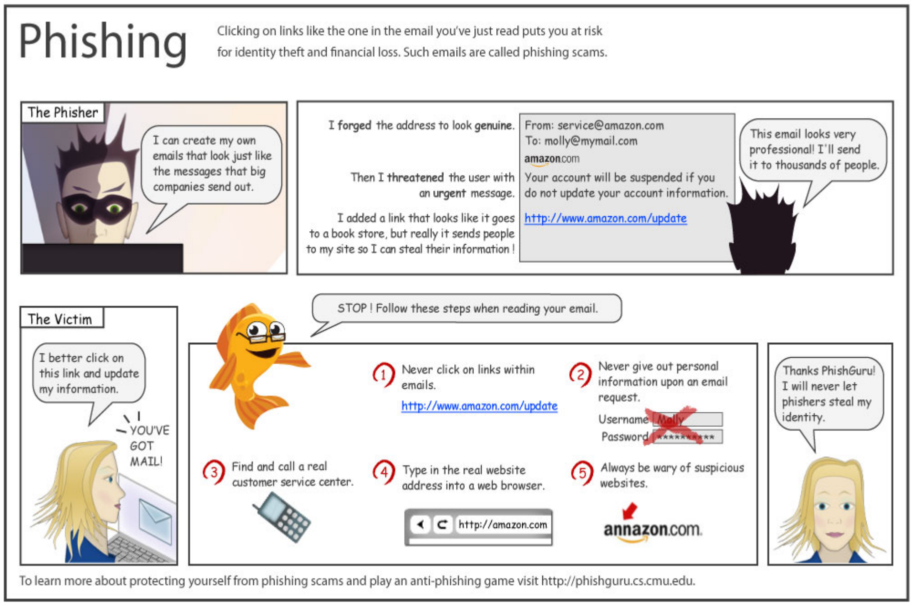

Phishing
Clicking on links like the one in the email you've just read puts you at risk for identity theft and financial loss. Such emails are called phishing scams.
The Phisher
I can create my own emails that look just like the messages that big companies send out.
I forged the address to look genuine.
Then I threatened the user with an urgent message.
I added a link that looks like it goes to a book store, but really it sends people to my site so I can steal their information!
From: service@amazon.com
To: molly@mymail.com
amazon.com
Your account will be suspended if you do not update your account information.
http://www.amazon.com/update
This email looks very professional! I'll send it to thousands of people.
The Victim
I better click on this link and update my information.
YOU'VE GOT MAIL!
STOP ! Follow these steps when reading your email.
1 Never click on links within emails.
http://www.amazon.com/update
2 Never give out personal information upon an email request.
Username Molly
Password **********
3 Find and call a real customer service center.
4 Type in the real website address into a web browser.
http://amazon.com
5 Always be wary of suspicious websites.
amazon.com
Thanks PhishGuru! I will never let phishers steal my identity.
To learn more about protecting yourself from phishing scams and play an anti-phishing game visit http://phishguru.cs.cmu.edu.

Figure 5.7: This is an update to Figure 5.6. We changed the PhishGuru character from a male to a fish (gender neutral). This design was used in the real world study [113].

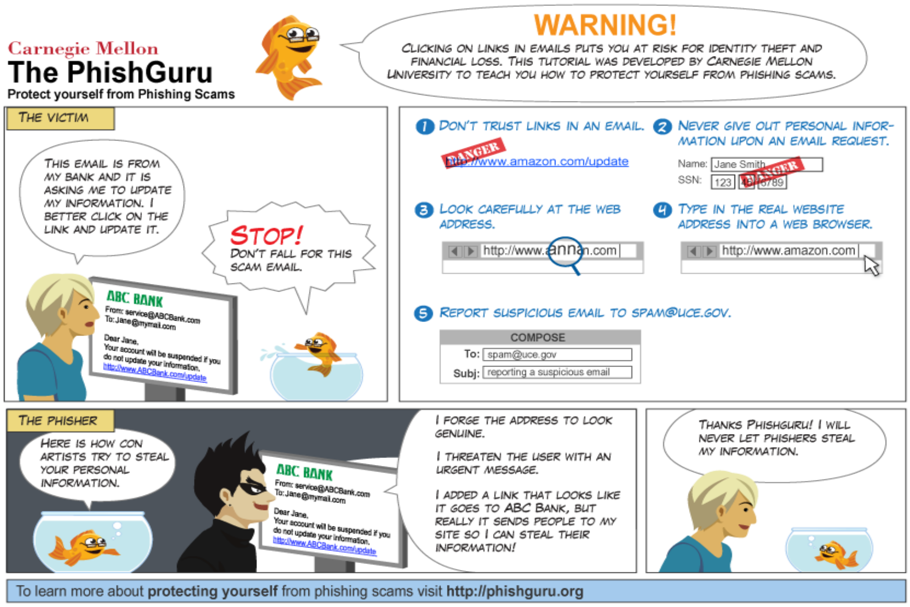

Carnegie Mellon
The PhishGuru
Protect yourself from Phishing Scams
WARNING!
CLICKING ON LINKS IN EMAILS PUTS YOU AT RISK FOR IDENTITY THEFT AND FINANCIAL LOSS. THIS TUTORIAL WAS DEVELOPED BY CARNEGIE MELLON UNIVERSITY TO TEACH YOU HOW TO PROTECT YOURSELF FROM PHISHING SCAMS.
THE VICTIM
THIS EMAIL IS FROM MY BANK AND IT IS ASKING ME TO UPDATE MY INFORMATION. I BETTER CLICK ON THE LINK AND UPDATE IT.
STOP! DON'T FALL FOR THIS SCAM EMAIL.
ABC BANK
From: service@ABCBank.com
To: Jane@mymail.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.ABCBank.com/update
1 DON'T TRUST LINKS IN AN EMAIL.
DANGER
http://www.amazon.com/update
2 NEVER GIVE OUT PERSONAL INFORMATION UPON AN EMAIL REQUEST.
Name: Jane Smith
SSN: 123 45 6789
DANGER
3 LOOK CAREFULLY AT THE WEB ADDRESS.
http://www.annan.com
4 TYPE IN THE REAL WEBSITE ADDRESS INTO A WEB BROWSER.
http://www.amazon.com
5 REPORT SUSPICIOUS EMAIL TO SPAM@UCE.GOV.
COMPOSE
To: spam@uce.gov
Subj: reporting a suspicious email
THE PHISHER
HERE IS HOW CON ARTISTS TRY TO STEAL YOUR PERSONAL INFORMATION.
I FORGE THE ADDRESS TO LOOK GENUINE.
I THREATEN THE USER WITH AN URGENT MESSAGE.
I ADDED A LINK THAT LOOKS LIKE IT GOES TO ABC BANK, BUT REALLY IT SENDS PEOPLE TO MY SITE SO I CAN STEAL THEIR INFORMATION!
THANKS PHISHGURU! I WILL NEVER LET PHISHERS STEAL MY INFORMATION.
To learn more about protecting yourself from phishing scams visit http://phishguru.org

Figure 5.8: This updated version of Figure 5.7 uses comic fonts and stresses the instructions. This version was used in the focus group studies we conducted.

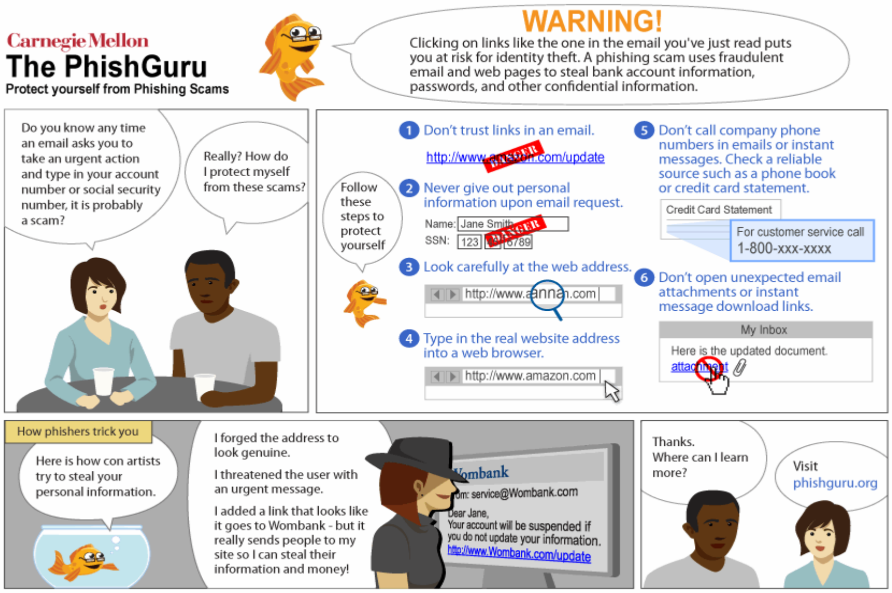

Carnegie Mellon
The PhishGuru
Protect yourself from Phishing Scams
WARNING!
Clicking on links like the one in the email you've just read puts you at risk for identity theft. A phishing scam uses fraudulent email and web pages to steal bank account information, passwords, and other confidential information.
Do you know any time an email asks you to take an urgent action and type in your account number or social security number, it is probably a scam?
Really? How do I protect myself from these scams?
Follow these steps to protect yourself
1 Don't trust links in an email.
http://www.[DANGER]n.com/update
2 Never give out personal information upon email request.
Name: Jane Smith
SSN: 123 DANGER 6789
3 Look carefully at the web address.
http://www.aannan.com
4 Type in the real website address into a web browser.
http://www.amazon.com
5 Don't call company phone numbers in emails or instant messages. Check a reliable source such as a phone book or credit card statement.
Credit Card Statement
For customer service call 1-800-xxx-xxxx
6 Don't open unexpected email attachments or instant message download links.
My Inbox
Here is the updated document.
attachment
How phishers trick you
Here is how con artists try to steal your personal information.
I forged the address to look genuine.
I threatened the user with an urgent message.
I added a link that looks like it goes to Wombank - but it really sends people to my site so I can steal their information and money!
Wombank
From: service@Wombank.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.Wombank.com/update
Thanks. Where can I learn more?
Visit phishguru.org

Figure 5.9: Updated version of Figure 5.1. This intervention was used in a large scale real world study [108].

To test the effectiveness of PhishGuru, we conducted a series of laboratory and real world studies which we describe in the next two chapters. PhishGuru is currently being commercialized by Wombat Security Technologies.[2]

[2] http://wombatsecurity.com/

Chapter 6

Laboratory Evaluation of PhishGuru

In this chapter, we present two laboratory studies we conducted to evaluate the effectiveness of PhishGuru. In Section 6.1, we present the first laboratory study, which compared the effectiveness of typical email security notices to the effectiveness of PhishGuru training [109]. In Section 6.2, we present a study measuring users' knowledge retention and knowledge transfer after PhishGuru training [110].

6.1 Preliminary evaluation of PhishGuru

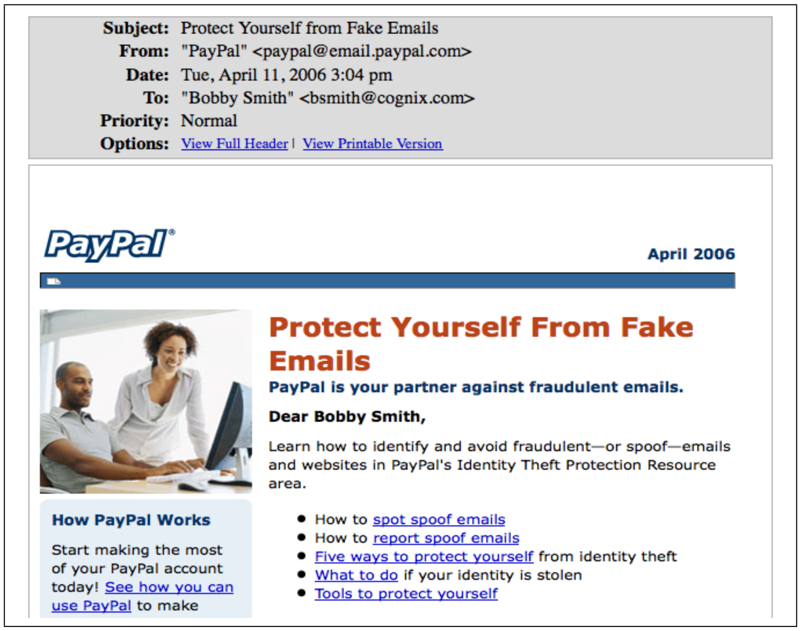
> This section is largely a reproduction of a paper co-authored with Yong Rhee, Alessandro Acquisti, Lorrie Cranor, Jason Hong, and Elizabeth Nunge and published at CHI 2007 [109].

In the first section of this chapter, we describe the evaluation of two PhishGuru interventions: one that provides a warning as well as actionable items using text and graphics, and one that uses a comic strip format to convey the same information. We present the results of a user study that compares the effectiveness of these two designs to the effectiveness of typical email security notices sent by e-commerce companies, who use these notices to alert their customers about phishing. In Section 6.1.1, we discuss participant recruitment and demographics; in Section 6.1.2, we present the hypotheses for the first laboratory study. In Section 6.1.3, we present the study methodology and the emails that we used as part of the study. In Section 6.1.4, we present the results of the evaluation, demonstrating that typical email security notices are ineffective, while embedded training designs are effective. In Section 6.1.5, we discuss some implications of the results.

6.1.1 Participant recruitment and demographics

As this research was focused on educating novice users about phishing attacks, we recruited participants with little technical knowledge. We posted fliers around the university and local neighborhoods, screening users through an online survey. We recruited users who said they had done no more than one of the following: changed preferences or settings in their web browser, created a web page, and helped someone fix a computer problem. In other studies, this approach has served as a good filter for recruiting non-experts [54, 107].

We recruited 30 particpants for the study, with each condition having 10 participants. Each participant was randomly placed in one of three groups. The "notices" group was shown the typical security notices displayed in Figure 6.1, while the "text/graphic" group was shown the text and graphics intervention displayed in Figure 6.2. The "comic" group was shown the comic strip intervention displayed in Figure 6.3. Table 6.1 shows the demographics of participants.[1]



Subject: Protect Yourself from Fake Emails
From: "PayPal" <paypal@email.paypal.com>
Date: Tue, April 11, 2006 3:04 pm
To: "Bobby Smith" <bsmith@cognix.com>
Priority: Normal
Options: View Full Header | View Printable Version

PayPal

April 2006

Protect Yourself From Fake Emails

PayPal is your partner against fraudulent emails.

Dear Bobby Smith,

Learn how to identify and avoid fraudulent—or spoof—emails and websites in PayPal's Identity Theft Protection Resource area.

- How to spot spoof emails
- How to report spoof emails
- Five ways to protect yourself from identity theft
- What to do if your identity is stolen
- Tools to protect yourself

How PayPal Works

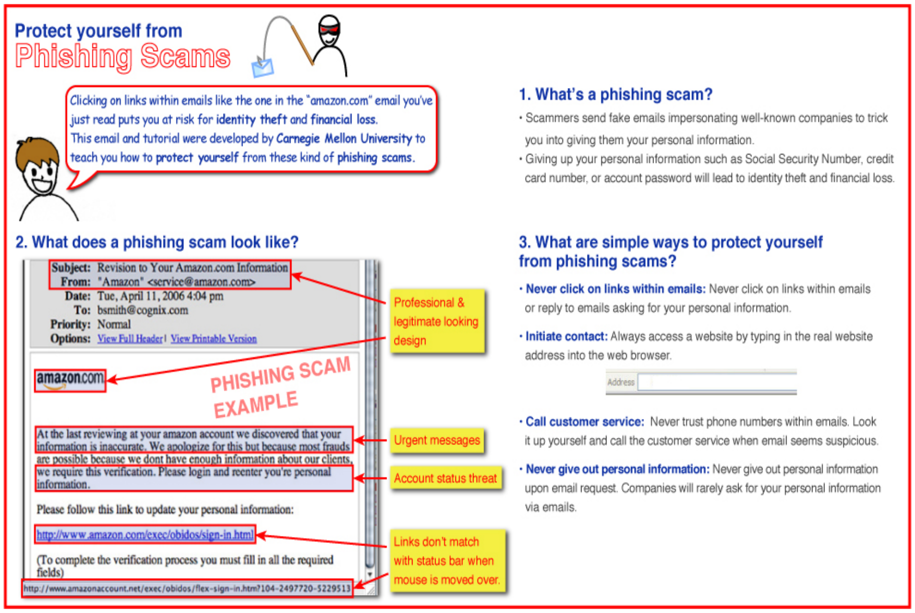
Start making the most of your PayPal account today! See how you can use PayPal to make

Figure 6.1: Security notices sent out by e-commerce companies to alert their customers about phishing.

6.1.2 Hypotheses

The two hypotheses that guided this study were:

[1] One outlier in the notices group received 300 emails daily, but did not perform particularly better or worse than others in this group. We found no significant relationship between propensity to fall for phishing attacks before the intervention and any of the demographic information we collected. Other small studies have also found no correlation between these demographics and susceptibility to phishing [53, 54].



Protect yourself from

Phishing Scams

Clicking on links within emails like the one in the "amazon.com" email you've just read puts you at risk for **identity theft** and **financial loss**. This email and tutorial were developed by **Carnegie Mellon University** to teach you how to **protect yourself** from these kind of **phishing scams**.

1. What's a phishing scam?

- Scammers send fake emails impersonating well-known companies to trick you into giving them your personal information.
- Giving up your personal information such as Social Security Number, credit card number, or account password will lead to identity theft and financial loss.

2. What does a phishing scam look like?

Subject: Revision to Your Amazon.com Information
From: "Amazon" <service@amazon.com>
Date: Tue, April 11, 2006 4:04 pm
To: bsmith@cognix.com
Priority: Normal
Options: View Full Header | View Printable Version

amazon.com

PHISHING SCAM EXAMPLE

At the last reviewing at your amazon account we discovered that your information is inaccurate. We apologize for this but because most frauds are possible because we dont have enough information about our clients, we require this verification. Please login and reenter you're personal information.

Please follow this link to update your personal information:

http://www.amazon.com/exec/obidos/sign-in.html

(To complete the verification process you must fill in all the required fields)

http://www.amazonaccount.net/exec/obidos/flex-sign-in.htm?104-2497720-5229513

Professional & legitimate looking design

Urgent messages

Account status threat

Links don't match with status bar when mouse is moved over.

3. What are simple ways to protect yourself from phishing scams?

- **Never click on links within emails:** Never click on links within emails or reply to emails asking for your personal information.
- **Initiate contact:** Always access a website by typing in the real website address into the web browser.
- **Call customer service:** Never trust phone numbers within emails. Look it up yourself and call the customer service when email seems suspicious.
- **Never give out personal information:** Never give out personal information upon email request. Companies will rarely ask for your personal information via emails.

Figure 6.2: The text/graphic intervention includes text with an annotated image of a sample phishing email.

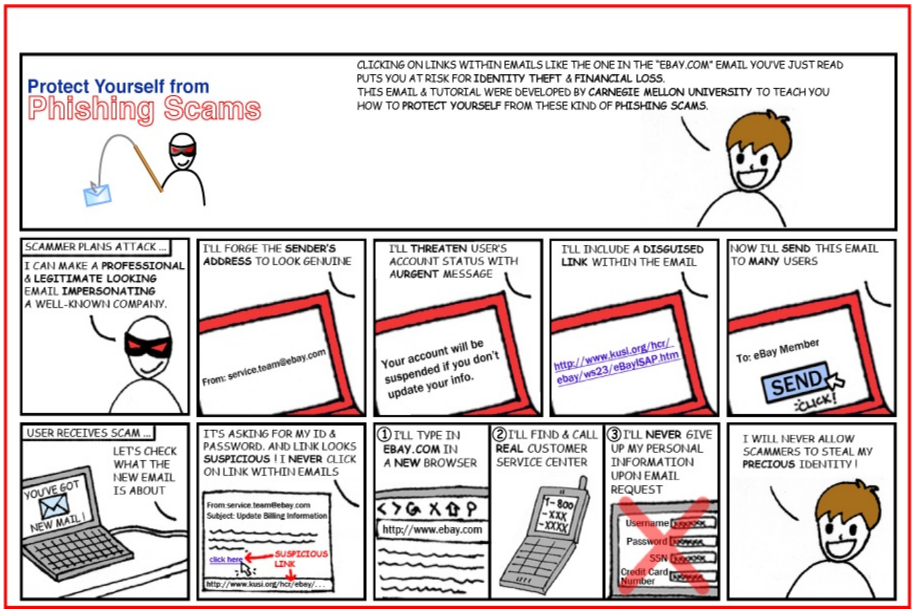

Protect Yourself from
Phishing Scams
CLICKING ON LINKS WITHIN EMAILS LIKE THE ONE IN THE "EBAY.COM" EMAIL YOU'VE JUST READ PUTS YOU AT RISK FOR IDENTITY THEFT & FINANCIAL LOSS. THIS EMAIL & TUTORIAL WERE DEVELOPED BY CARNEGIE MELLON UNIVERSITY TO TEACH YOU HOW TO PROTECT YOURSELF FROM THESE KIND OF PHISHING SCAMS.
SCAMMER PLANS ATTACK ...
I CAN MAKE A PROFESSIONAL & LEGITIMATE LOOKING EMAIL IMPERSONATING A WELL-KNOWN COMPANY.
I'LL FORGE THE SENDER'S ADDRESS TO LOOK GENUINE
From: service.team@ebay.com
I'LL THREATEN USER'S ACCOUNT STATUS WITH AURGENT MESSAGE
Your account will be suspended if you don't update your info.
I'LL INCLUDE A DISGUISED LINK WITHIN THE EMAIL
http://www.kusi.org/hcr/ebay/ws23/eBayISAP.htm
NOW I'LL SEND THIS EMAIL TO MANY USERS
To: eBay Member
SEND
CLICK!
USER RECEIVES SCAM ...
LET'S CHECK WHAT THE NEW EMAIL IS ABOUT
YOU'VE GOT NEW MAIL!
IT'S ASKING FOR MY ID & PASSWORD. AND LINK LOOKS SUSPICIOUS ! I NEVER CLICK ON LINK WITHIN EMAILS
From:service.team@ebay.com
Subject: Update Billing Information
click here
SUSPICIOUS LINK
http://www.kusi.org/hcr/ebay/...
① I'LL TYPE IN EBAY.COM IN A NEW BROWSER
http://www.ebay.com
② I'LL FIND & CALL REAL CUSTOMER SERVICE CENTER
1-800-XXX-XXXX
③ I'LL NEVER GIVE UP MY PERSONAL INFORMATION UPON EMAIL REQUEST
Username
Password
SSN
Credit Card Number
I WILL NEVER ALLOW SCAMMERS TO STEAL MY PRECIOUS IDENTITY !

Figure 6.3: The comic strip intervention uses a comic strip to tell a story about how phishing works and how people can protect themselves.

Table 6.1: Study participants

Characteristics	Notices	Text/ Graphic	Comic
Sample size	10	10	10
Gender			
Male	50%	40%	20%
Female	50%	60%	80%
Browser			
IE	80%	60%	60%
Firefox	10%	20%	30%
Others	10%	20%	10%
Average emails per day	51.4	36.9	15
Average age in years	31.2	27.5	21.1

Hypothesis 1: Security notices are an ineffective medium to teach users about phishing attacks.

Hypothesis 2: Participants in the embedded training group (text/graphics and comic) will learn more effectively than participants in the security notices group.

6.1.3 Methodology

The laboratory study consisted of a think-aloud session in which participants played the role of "Bobby Smith," an employee of Cognix Inc. who works in the marketing department. Participants were told that the study investigated "how people effectively manage and use emails." They were told to interact with their email the way they normally do in real life. If a participant was not familiar with SquirrelMail (a web-based email client), we gave that participant a quick tutorial describing how to perform simple actions. Participants used a 1.40GHz Compaq laptop running Microsoft Windows XP home edition with Internet Explorer 6.0 to access emails. We also mentioned that we would be able to answer questions about using SquirrelMail during the study, but would not be able to help them make any decisions. We asked participants a few pre-study questions about their use of email to reinforce the idea that this was a study about the use of email systems. We recorded audio and screen interactions using Camtasia.

We gave participants an information sheet describing the scenario and asked them to read it aloud and ask clarification questions. The information sheet included the usernames and passwords for Bobby Smith's email account and accounts at Amazon, American Express, Citibank, eBay and PayPal. We also provided username and password information in a physical wallet that participants could use throughout the study.

Each participant was shown 19 email messages arranged in the predefined order shown in Table 6.2. Nine *legitimate-no-link* messages were legitimate emails without any links, received from co-workers at Cognix, friends, and family. These emails requested that Bobby Smith perform simple tasks like checking the status of products at Staples. Two *legitimate-link* messages were simulated legitimate emails from organizations with which Bobby Smith had an account. The mailbox also contained two *spam* emails, two *phishing-account* fraudulent emails that appeared to come from organizations where Bobby had an account, and two *phishing-no-account* fraudulent emails that appeared to come from a bank with which Bobby did not have an account. The mailbox also had two *training* emails—security notices or embedded training interventions depending on the group to which the participant belonged. Table 6.3 presents examples of each type of email we used in the study.

Table 6.2: Email arrangement in the study.

1. Legitimate-no-link
2. Legitimate-no-link
3. Phishing-account
4. Legitimate-no-link
5. Training
6. Legitimate-no-link
7. Legitimate-link-account
8. Spam-link-no-account
9. Legitimate-no-link
10.Legitimate-no-link
11. Training
12. Spam-link-no-account
13. Legitimate-link-no-account
14. Phishing-no-account
15. Legitimate-no-link
16. Phishing-no-account
17. Phishing-account
18. Legitimate-no-link
19. Legitimate-no-link

Table 6.3: Sample of emails used in the first laboratory study.

Email type	**Sender information**	**Email subject line**
Legitimate-no-link	Brandy Anderson	Booking hotel rooms for visitors
Legitimate-link	Joseph Dicosta	To check the status of the product on Staples
Phishing-no-account	Wells Fargo	Update your bank account information!
Phishing-account	PayPal	Reactivate you PayPal account!
Spam	Eddie Arredondo	Fw: Re: You will want this job
Training	Amazon	Revision to your Amazon.com information

All of the phishing, spam, and security notice emails we used for this study were based on actual emails we had collected. We created exact replicas of the phishing websites on a local machine by running Apache and modifying the host files in Windows so that IE would display the URL of the actual phishing websites. All replicated phishing websites were completely functional and allowed participants to submit information.

We used a completely functional SquirrelMail implementation to allow users to access Bobby Smith's email. We wrote a Perl script to push emails into the SquirrelMail server; this script was also used to change the training emails for each group.

After participants finished going through Bobby Smith's emails, we asked them some post-study questions and debriefed them. In the debriefing, we asked them questions about their choices during the study. We also showed them training messages belonging to a different group than the one they had been placed in for the study. For example, participants who viewed Figure 6.2 in their study were shown Figure 6.3 after the study and vice versa. They were then asked for their views on both designs.

6.1.4 Results

In this section we present the results of the first laboratory study. In the analysis, we considered someone to have fallen for a phishing attack if they clicked on a link in a phishing email, regardless of whether they went on to provide personal information. Although not everyone who clicks on a phishing link will go on to provide personal information to a website, people in this study who clicked on phishing links provided information 93% of the time. What's more, clicking on phishing links can be dangerous even if someone does not actually provide personal information to the site because some phishing sites transmit malware to a user's computer.

Security Notices Intervention

There was no difference between the number of participants who clicked on links in phishing emails before and the number who clicked after the two security notice messages. The first security notice users saw was one of two security messages that eBay or PayPal sends to their customers. This notice email was linked to a real website [56]. Only five (50%) users in this group clicked on the first security notice link in the email to learn more about phishing attacks. Among these five participants, only two (40%) actually read through the content in the web pages, while the other three (60%) skimmed through the content and closed the window. Nine (90%) participants clicked on the second security notice; this security notice was sent from the system administrator of Cognix. During the post-study debriefing, we asked whether the notices had been helpful. The participants who had seen the security notices said the information took too long to read and that they were not sure what the messages were trying to convey. Nine participants (90%) fell for the phishing email sent before the security notice email and nine participants (90%) fell for the final phishing email. The mean percentage of participants who fell for the three phishing emails presented after the security notices was 63%.

Text/graphics Intervention

In this group, eight participants (80%) fell for the first phishing email and all ten participants clicked on the training message link in the training email. Seven participants (70%) clicked on the second training message and seven participants (70%) fell for the final phishing email. The mean percentage of participants who fell for the three phishing emails presented after the interventions was 30%. After going through the training message, many participants checked to see if they had an account with the relevant financial institution before clicking on the link. Only one user (10%) clicked on the phishing message that was sent from Barclays Bank, which they did not have an account with. When asked why she had done so, the user said, "just because it [the link] was there and I wanted to check what they show."

Most participants liked the way the information was presented; a common comment was: "Having the image and the text with callouts was helpful." One user said: "Giving the steps to follow to protect from phishing was helpful." Another mentioned, "This is definitely useful and good stuff and I will remember that [to look for URLs in the status bar]."

Comic Strip Intervention

The results of this study indicate that the comic strip intervention was most effective in educating people about phishing attacks. All of the participants in this group fell for the first phishing email and also clicked on the training message. Six participants (60%) clicked on the second training message; only three participants (30%) fell for the final phishing email. The mean percentage of participants falling for the three phishing emails presented after the interventions was 23%. Some participants said they preferred the comic to the text/graphics intervention because it engaged them with a story. However, other participants felt that the text/graphics version was more serious and professional. One user said, "The comic version is good for children but I would prefer text with the image."

Comparison

The security notices group and the comic group displayed significantly different levels of ability to recognize phishing emails. In the notices group, nine participants (90%) fell for the final phishing email; in the comic group, only 3 participants (30%) fell for this email (Chi-Sq = 23.062, DF = 1, p-value = 0.001).

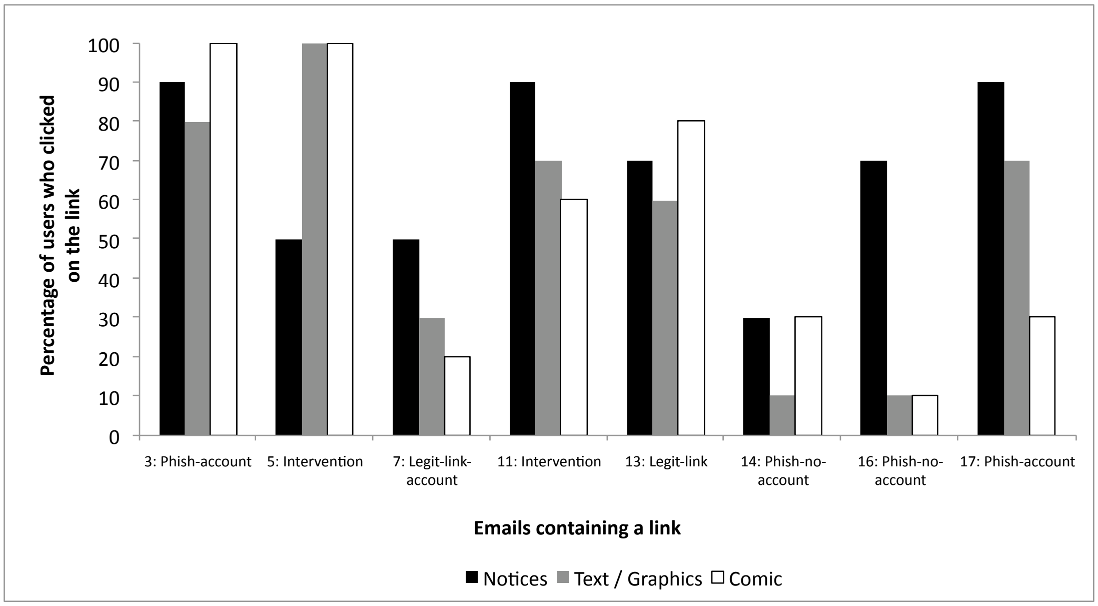


Figure 6.4: Comparison of different methods of training for each group for all of the emails which had a link in them. The number represents the location of the email in the email arrangement. Participants in the Comic strip group were able to identify phishing emails better than the other two groups.

We also compared the effectiveness of security notices to the effectiveness of the text and graphic intervention. The number of participants who fell for phishing attacks before and after training in the notices group was nine (90%), while the number of participants falling for phishing attacks in the text/graphics group was eight (80%) before training and seven (70%) after training. The difference between these two groups was not as significant (Chi-Sq = 0.364, DF = 1, p-value = 0.546) as the difference between the notices and comic groups.

There was a significant difference in effectiveness between the two embedded training interventions (Chi-Sq = 16.880, DF = 1, p-value = 0.001). The mean scores across the three phishing emails after intervention was lowest for the comic group. Figure 6.4 presents a comparison of the three training methodologies for all of the emails containing links.

In the post-study questions we asked participants from the comic and text/graphics groups the following question: "Which one [design] would you prefer and why would you prefer it?" Nine (45%) of the twenty participants preferred the comic version of the information representation and eleven (55%) preferred the text with graphics version.

During the post-study session, we asked specific questions about the training methodology and how these methods raised phishing awareness. One of the questions was: "Did the method create awareness about phishing attacks?" Only two (20%) participants said the security notices raised awareness about phishing attacks, while in both of the other groups all participants (100%) said the method they encountered raised awareness about phishing attacks. We also asked participants:

“Do you think this method will help you learn techniques to identify false websites and email?” None of the participants said the security notices would help them, while all of the participants in the other groups thought the embedded training messages would help them.

We also compared data for the individual performance of the participants before and after training. We observed that 9 out of 10 participants (90%) in the notices group clicked the first phishing email and that of these, 8 participants (89%) clicked on the final phishing email. In the text/graphics group, 8 participants (80%) clicked on the first phishing email, out of which 5 (63%) clicked on the final phishing email. In the comic group, 10 participants (100%) clicked on the first phishing email, out of which 3 participants (30%) clicked on the final phishing email. We found that participants in the security notices group performed significantly differently from participants in the comic group (Chi-Sq = 18.245, DF = 1, p-value = 0.001). There was also a significant difference between the performances of participants in the text/graphics group and those in the comic group (Chi-Sq = 7.222, DF = 1, p-value = 0.007). There was no significant difference between the performance of participants in the notices group and those in the text/graphics group.

During the post-study session, we also asked the participants: “On a scale of 1 to 7, where 1 is not at all confident and 7 is most confident, how confident were you while making decisions on clicking links and replying to emails?” In the notices group, the values ranged from 4 to 7 (mean = 5.4, s.d. = 1.1, variance = 1.2); in the text/graphics group, values ranged from 3 to 6 (mean = 4.6, s.d. = 0.9, variance = 0.8); in the comic group, values ranged from 3 to 7 (mean = 5.5., s.d. = 1.3, variance = 1.6). Participants in the three groups had similar levels of confidence in handling emails.

General observations

Participants seemed to identify the Nigerian scam email (email number 12) easily. Only two of the thirty participants (6.7%) clicked on the link in this email. Only nine participants (30%) actually clicked on the link in the second phishing email (email number 14), which was ostensibly from a company they did not have an account with. Among these nine participants, four (44.4%) realized that they did not have an account with the service once they clicked on the link; as a result, they closed the window immediately.

Twenty-four (80%) of the participants were not familiar with the mouse-over technique, which can be used to view the actual URL before clicking on a link. Most participants appreciated being taught such a technique. One user said, “I did not know to look for links before [in email], I will do it now.”

One user in the text/graphics group did not click on any links in the emails because of her personal experience as a victim of identity theft. This user stated, “I was a victim of online credit card fraud, so from then on I decided not to click on links in the emails." No user in the study actually entered random information to test the phishing site's reaction. Two participants used search engines to help decide how to react to an email. One user Googled the phrase "Bank of Africa" from the Nigerian scam. Another user said, "I will ask one of my friends to help me make a decision here, she knows about these things better than me."

Among participants who did not understand the training messages, we observed behavior similar to that discussed by Dhamija et al. [53]. Novice users used misleading signals [107] to make their decisions. For example, one of the participants used the lock icon on the phishing website we created to decide that the website was legitimate. When asked why she did that, she said: "I do that often to find whether the website is legitimate." Another participant mentioned that "the logo [Citibank] is real so the site must be legitimate." Another participant said, "I visited this website [PayPal] some days back. It looks the same as before, so it must be legitimate." A few other participants were satisfied that the website was legitimate because it displayed updated account information after they entered their personal information.

Repetitive training in a short time span was helpful for some participants. Some participants did not understand what was going on the first time the training information was presented, but read it carefully the second time. To study repetitive training, we sent two training messages separated by 14 days in one of our real-world studies (discussed in Section 7.2). We found that adding a second training message to reinforce the original training decreases the likelihood of people giving information to phishing websites.

6.1.5 Discussion

We conducted lab experiments comparing the effectiveness of two interventions with standard security notices about phishing. Results from this study supported Hypothesis 1 and Hypothesis 2 introduced in Section 6.1.2.

As observed in other studies, we saw that novice users used misleading signals to make decisions. We believe that properly designed training messages and interventions can help novice users detect and use meaningful signals.

These results strongly suggest that security notices fail to effectively teach people about phishing attacks. We believe this is because people do not understand why they are receiving such emails, and also because it is difficult for them to relate to an abstract problem they may not believe is likely to occur. In addition, some participants claimed that they knew about phishing and knew how to protect themselves, but ultimately fell for the phishing scams anyway. This also suggests that people may be overconfident about what they know, especially if they have seen such security notices in the past, and thus disregard new notices when they appear.

The results also indicate that the comic strip intervention was most effective. The primary differences between the two interventions was that the comic strip format had significantly less text and more graphics, and that it told a story to convey its message.

6.2 Evaluation of retention and transfer

> This section is largely a reproduction of a paper co-authored with Yong Rhee, Steve Sheng, Sharique Hasan, Alessandro Acquisti, Lorrie Cranor, and Jason Hong and published at e-Crime Researchers Summit 2007 [110].

In the study described in Section 6.1, we tested users immediately after training and demonstrated that embedded training improved users' ability to identify phishing emails and websites. We also compared embedded training to security notices delivered via email. In this section, we present a study in which we tested users to determine how well they retained knowledge gained through embedded training over a period of about one week; we also tested how well they used this knowledge to identify other types of phishing emails. We further compared the effectiveness of training materials delivered via embedded training to those delivered as a regular email message (non-embedded).

In Section 6.2.1, we present the theory and the hypotheses that guided the study. In Section 6.2.2, we present the participant recruitment methodology and demographics. In Section 6.2.3, we present the study methodology used to test the hypotheses. In Section 6.2.4, we present the results of the evaluation, demonstrating that embedded training is more effective than non-embedded training, and that users can retain learned information over time, transferring the knowledge they gained. We discuss the effect of training users in Section 6.2.5.

6.2.1 Theory and hypotheses

In this section we introduce five hypotheses for the following study. Three hypotheses relate to user learning, when two relate to users' susceptibility to phishing emails.

Learning

Motivation is one of the most important aspects of the learning process. Researchers have shown that users can be trained through an embedded training method that makes training part of the primary task. With this form of training, users are motivated to learn because they are presented with training materials immediately after they fall for phishing emails [109]. However, while the earlier study (Section 6.1) suggested that embedded training motivates people to learn, it did not evaluate whether the embedded training approach was better than sending the same training materials directly via email.

In the second laboratory study, we had four conditions: “embedded,” “non-embedded,” “suspicion,” and “control.” Participants in the embedded condition received a simulated phishing email and saw a revised version of the comic strip intervention when they clicked on a link in that email. Participants in the non-embedded condition received the same training materials directly as part of an email message; they did not have to fall for a simulated phishing email to see the training materials. Participants in the suspicion condition received a brief email from a friend that mentioned phishing without providing any information about how they could protect themselves.

> **Hypothesis 1**: Participants in the embedded condition learn more effectively than participants in the non-embedded condition, suspicion condition, and control condition.

Retention

A large body of literature focuses on quantifying knowledge retention [166]. Learning science literature defines retention as the ability of learners to retain or recall the concepts and procedures taught when tested under the same or similar situations after a time period δ from the time knowledge was acquired. Researchers have frequently debated the optimum δ to measure retention [134]. The first laboratory study that demonstrated that users can be taught to avoid phishing attacks tested users immediately after they were trained; as a result, it did not explore users' ability to retain this knowledge [110, 181]. Thus, the question remained as to whether users retain the knowledge they have gained during training.

> **Hypothesis 2**: Participants in the embedded condition retain more knowledge about how to avoid phishing attacks than participants in the non-embedded condition, suspicion condition, and control condition.

Transfer

Transfer is the ability to apply the knowledge gained from one situation to another situation after a time period δ from the time of knowledge acquisition. Researchers have emphasized that transferability of learning is of prime importance in training. Two types of transfers are discussed in the literature: *near transfer*, in which the testing situation is similar to the training situation, and *far transfer*, in which the testing situation is very different from the training situation [198]. In this study, we focused on measuring near transfer. For example, we trained users with an email regarding revision to their Amazon account and tested them with an email from PayPal regarding reactivation of their PayPal account.

Hypothesis 3: Participants in the embedded condition transfer more knowledge about how to avoid phishing attacks than participants in the non-embedded condition, suspicion condition, and control condition.

Cognitive reflection

Many user studies examining phishing or phishing-related interventions have been agnostic to individual user characteristics (sex, age, education level, and hours using the computer). Others have failed to find significant relationships between features such as age or gender and phishing-related behavior [53, 54, 109, 181]. This may be the product of one or more of the following factors: (1) individual differences (sex, age, education level, and hours using the computer) are not actually relevant to phishing-related behavior; (2) the sample sizes used for these studies were too small to detect any significant relationships; and (3) truly discriminating characteristics have not yet been tested. In this study, we not only retested previously studied demographic characteristics, but also investigated whether an individual's propensity for *cognitive reflection* related to their ability to avoid falling for phishing attacks.

People vary along many dimensions, and these variations often result in differences in behavior and decision-making. Frederick suggests that individuals who are more cognitively reflective differ from those who are less reflective [71]. To evaluate this characteristic, he presents the Cognitive Reflection Test (CRT), which consists of three questions whose correct solutions require the suppression of impulsivity. In his study, Frederick tested the CRT among approximately 3500 individuals at various universities and in several web-based studies. Although his three-question CRT does correlate highly with other measures of achievement and intelligence such as the Scholastic Aptitude Test (SAT) and the Wonderlic Personnel Test (WPT), Frederick argues that the CRT more accurately measures "cognitive reflection" or "the ability or disposition to resist reporting the response that first comes to mind." He found that higher CRT scores correlate with more risk-taking and lower discount rates. Conversely, those who are less cognitively reflective are more likely to choose certain gains over higher expected values and choose lower amounts immediately over larger rewards later.

The three questions included in the CRT are:

1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? _______ cents
2. If it takes five machines 5 minutes to make five widgets, how long would it take 100 machines to make 100 widgets? _______ minutes
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _______ days

The correct answers are: (1) 5 cents, (2) 5 minutes , and (3) 47 days.

With Cognitive Reflection as the measure of individual variation, we propose two hypotheses about the differential phishing-related behavior of the users. The first hypothesis draws from the idea that high CRT scores are associated with less impulsive behavior. This hypothesis suggests that individuals with high CRT scores will more thoroughly deliberate emails for which they have a mental model. We hypothesize the following:

> **Hypothesis 4**: Users with higher scores on the Cognitive Reflection Test (CRT) will be less likely than users with lower scores to click on phishing emails from companies with which they have an account.

On the other hand, the emails ostensibly sent from companies with which a user does not have an account (no-account) are not part of the user's mental model. In this situation, we predict that those with lower CRT scores will be less likely to deviate from the rules and thus will not click on the links in the no-account emails. On the other hand, we hypothesize that those with a higher CRT score, whom we expect to be greater risk-takers, will explore the no-account emails out of curiosity:

> **Hypothesis 5**: Confronted with a novel situation, those with higher scores on the CRT will be more likely than users with lower scores to click on the links in the phishing emails from companies with which they have no account.

6.2.2 Participant recruitment and demographics

We recruited participants by posting fliers in and around the university campus advertising an "email management study." We asked all respondents to complete an online screening survey. We selected people who did not know what phishing was and who had never taken part in any of our previous studies. The screening survey included questions like "What does the term cookie mean?" and "Approximately how many times have you used online banking services in the last 6 months?" so that people were not primed towards the idea that we might be conducting a phishing study.

The screening survey was filled out by 185 people; 87 (47%) people qualified for the study. Before administering the actual study, we conducted pilot studies with nine qualified participants. The pilot studies were used to refine the study methodology. Sixty-three of the 87 qualified people completed the actual study. However, the data from some participants was excluded from subsequent analysis because they had not viewed the training intervention. We thus analyzed data for 56 participants who had been randomly assigned to one of four conditions. Table 6.4 provides the demographic characteristics of the 56 participants whose data we analyzed.

Table 6.4: Demographics of the participants; N = 14 in each condition; the standard deviation is presented in parentheses.

Characteristics	Control	Suspicion	Embedded	Non-Embedded
Sample size	14	14	14	14
Gender				
Male	36%	50%	36%	43%
Female	64%	50%	64%	57%
Browser				
IE	64%	36%	50%	50%
Firefox	29%	57%	29%	43%
Others	7%	7%	21%	7%
Average emails per day	20.5	17.6	16.1	20.7
Average age in years	28	26.9	24.6	24.3
Average CRT score	1.3 (1.2)	1.1 (0.89)	1.25 (0.91)	1.14 (0.94)
Average time reading the intervention			97 seconds (32.5)	37 seconds (66.2)

6.2.3 Methodology

This study was conducted in two laboratory sessions separated by at least 7 days (mean = 7.2, s.d = 0.81). Participants came to the laboratory for a study investigating "how people effectively manage and use email." When they arrived to the laboratory for the first session, we had them fill out the pre-study questionnaire, which included demographic information along with the CRT questions.

The study consisted of two think-aloud sessions in which the participants played the role of "Bobby Smith," the business administrator for Cognix Inc. We had participants sit at a desk in the laboratory, which we told them was Bobby's office desk. The desk was outfitted with a laptop, pens, note pads, post-it notes, and other office supplies. Figure 6.5 shows the laboratory setup where we conducted the study. We provided the participants with a printout that included details about their role, including the names of people Bobby interacts with (co-workers, family, and friends) and all of the organizations where Bobby has an account. We also provided the participants with a printout of the user names and passwords for all of Bobby's accounts: AOL, Amazon, American Express, Bank of America, Citibank, eBay, Gmail, PayPal, Staples, and Yahoo. We showed each participant Bobby's email inbox and asked them to process the email and react to the email as they would in the real world, keeping in mind the role they were playing. When participants clicked on the links in fake phishing emails, they were presented with the PhishGuru interventions.

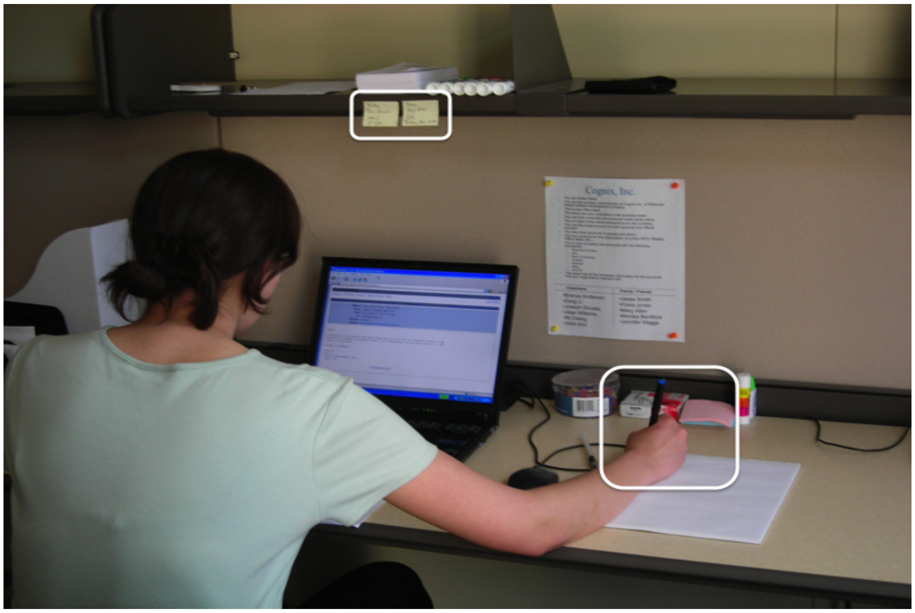


Figure 6.5: One of the participants playing the role of Bobby Smith. The top highlighted box shows the post-it notes that this participant made notes on and stuck to the bookshelf during the user study. The bottom highlighted box shows the participant taking additional notes on the notepad.

Participants saw the training intervention only in session 1. When participants completed session 1 of the study, they were not provided with any additional information about phishing or the nature of the study.

When participants came back after approximately seven days for the second session, we told them that they would be role-playing as Bobby Smith again, just as they had in session 1. Once again, we showed them Bobby's email inbox and asked them to process Bobby's email. We asked all participants at the end of session 2 to complete a post-study survey after they completed their email management tasks. We also de-briefed them about the study after they completed the post-study survey.

We used a 1.70GHz IBM T42 ThinkPad laptop running Microsoft Windows XP Home Edition to conduct the user studies. The participants used Internet Explorer 6.0 to access emails through SquirrelMail. We wrote a Perl script to push emails into the SquirrelMail server; this script was also used to setup Bobby's inbox for each participant. We recorded the participants' voices and screen-captured their interactions using Camtasia.

We designed the emails in Bobby's inbox to allow us to measure the immediate effectiveness of the interventions as well as knowledge retention and transfer. In session 1, participants saw 33 emails in Bobby's inbox: a set of 16 before-training emails (the "before" set), a training intervention, and a set of 16 additional emails shown immediately after training (the "immediate" set). In session 2, participants saw another 16 emails (the "delay" set) in Bobby's inbox. We had three sets of 16 emails (A, B, and C) that we used for the before, immediate, and delay sets. Each set consisted of 9 legitimate emails without any links in them from people with whom Bobby interacts (legitimate-no-link), 3 legitimate emails containing links from organizations and people with whom Bobby interacts (legitimate-link), 2 phishing emails from organizations where Bobby has an account (phishing-account), 1 email from a bank with which Bobby does not have an account (phishing-no-account), and 1 spam email. Participants were randomly assigned to see either set A or set C as the before set and the other one as the delay set. All participants saw set B as the immediate set. Table 6.5 summarizes the contents of email set A. Sets B and C contained the same types of emails with a different combination of senders and subjects.

All participants in the embedded and non-embedded training conditions saw a training intervention from Amazon, a company with which Bobby has an account, with the subject "Revision to your Amazon.com information." Participants in the embedded condition saw the training material shown in Figure 6.6 when they clicked on the link in the email, while those in the non-embedded condition received the training message in the email itself. Participants in the suspicion condition did not receiving any training material, instead received an email from a friend.

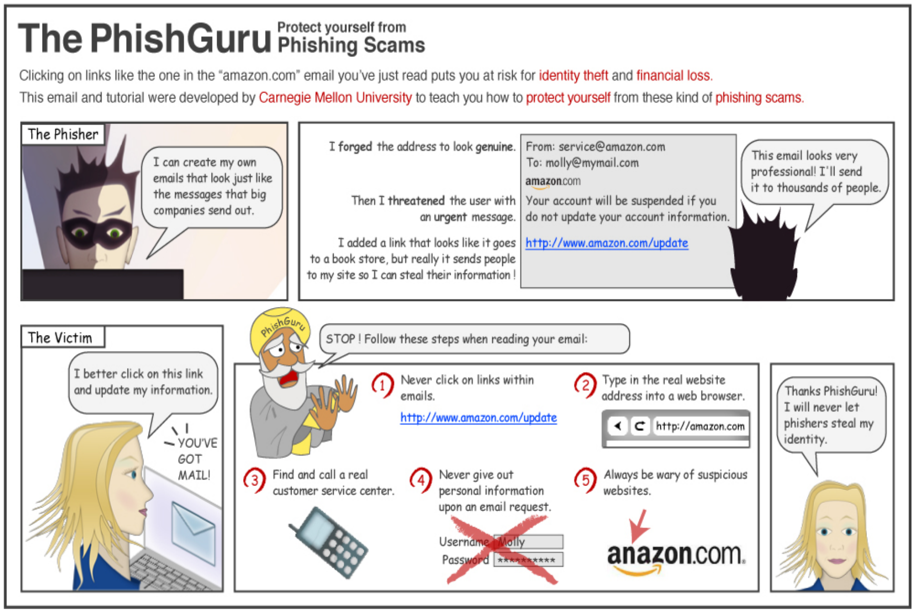

The PhishGuru
Protect yourself from Phishing Scams
Clicking on links like the one in the "amazon.com" email you've just read puts you at risk for identity theft and financial loss.
This email and tutorial were developed by Carnegie Mellon University to teach you how to protect yourself from these kind of phishing scams.
The Phisher
I can create my own emails that look just like the messages that big companies send out.
I forged the address to look genuine.
Then I threatened the user with an urgent message.
I added a link that looks like it goes to a book store, but really it sends people to my site so I can steal their information !
From: service@amazon.com
To: molly@mymail.com
amazon.com
Your account will be suspended if you do not update your account information.
http://www.amazon.com/update
This email looks very professional! I'll send it to thousands of people.
The Victim
I better click on this link and update my information.
YOU'VE GOT MAIL!
PhishGuru
STOP ! Follow these steps when reading your email:
1 Never click on links within emails.
http://www.amazon.com/update
2 Type in the real website address into a web browser.
http://amazon.com
3 Find and call a real customer service center.
4 Never give out personal information upon an email request.
Username Molly
Password **********
5 Always be wary of suspicious websites.
anazon.com
Thanks PhishGuru! I will never let phishers steal my identity.

Figure 6.6: Revised comic strip intervention design. The top row presents the activities of a phisher while the bottom row shows the victim and presents steps the victim can take to avoid falling for the phishing attack.

All of the phishing, spam, and legitimate-with-link emails used for this study were based on actual emails collected from members of the research group. We designed the legitimate-no-link emails to resemble emails that one of the business administrators at the University typically receives. We created exact replicas of the phishing websites on the local machine by running Apache and modifying the host files in Windows so that IE would display the URL of the actual phishing websites. All replicated phishing websites were completely functional and allowed people to submit information. These phishing websites were only accessible from the laboratory machine used for the user studies. Users were taken to these phishing websites when they clicked on links in the phishing-account and phishing-no-account emails.

6.2.4 Results

In this section, we present the results of the user study we conducted to test the five hypotheses introduced in Section 6.2.1. We consider someone to have fallen for a phishing attack if they click on a link in a phishing email, regardless of whether they go on to provide personal information. The conclusions presented in this section are robust to the selection of a different metric for the evaluation of the correctness of participants' choices. Specifically, the findings listed in this section persist when the participants provide personal information to a phishing website during the experiment, rather than simply clicking on the links in a spoofed email. Although not everyone who clicks on a phishing link will go on to provide personal information to a website, in this study people who clicked on phishing links provided information 90% of the time. We calculated correctness scores as the number of emails containing links that a participant correctly identified as phishing or legitimate. We determined the correctness of that identification based on whether or not the participant clicked on a link in each email.

The results of the study supported hypotheses 1, 2, 3, and 5; they rejected hypothesis 4. We found no correlation between participants' scores (correctly identifying phishing emails as phishing and legitimate emails as legitimate) and participants' demographics. We found that, after the training, participants in the embedded condition made better decisions than participants in the non-embedded condition. In fact, participants in the non-embedded condition did not perform significantly better after training than those in the control condition (who had received no training). Also, participants in the embedded condition spent significantly more time reading the intervention than participants in the non-embedded condition. We found that participants in the embedded condition retained and transferred more knowledge than participants in the non-embedded condition. We also found that participants with higher Cognitive Reflection Test (CRT) scores were more likely than users with lower CRT scores to click on the links in the phishing emails from companies with whom they did not have a account. Furthermore, we found that participants generally liked the embedded training methodology and intervention design (comic strip) used for the study.

Table 6.5: Arrangement of email in set A. The other sets had similar distribution of emails.

#	Email subject line information	legitimate-no-link	legitimate-link	phishing-account	phishing-no-account	Spam
1	[cognix] REMINDER: Dont forget to attend the tax session	√				
2	RE: Room booking - Sunday - To meet - Let me know	√				
3	Reactivate you PayPal account!			√		
4	Booking hotel rooms for visitors	√				
5	Re: Funny joke (fwd)	√				
6	Fw: Re: You will want this job					√
7	To check the status of the product on Staples		√			
8	Dont forget moms birthday!	√				
9	Update your bank account information!				√	
10	Please check PayPal balance		√			
11	coffee from starbucks	√				
12	RE: Tea powder - Kitchen	√				
13	IMPORTANT: Please Update Your AOL account			√		
14	New member in our administrative team	√				
15	Confirmation: Payment Received		√			
16	Sorry missed your call - will call you this weekend	√				

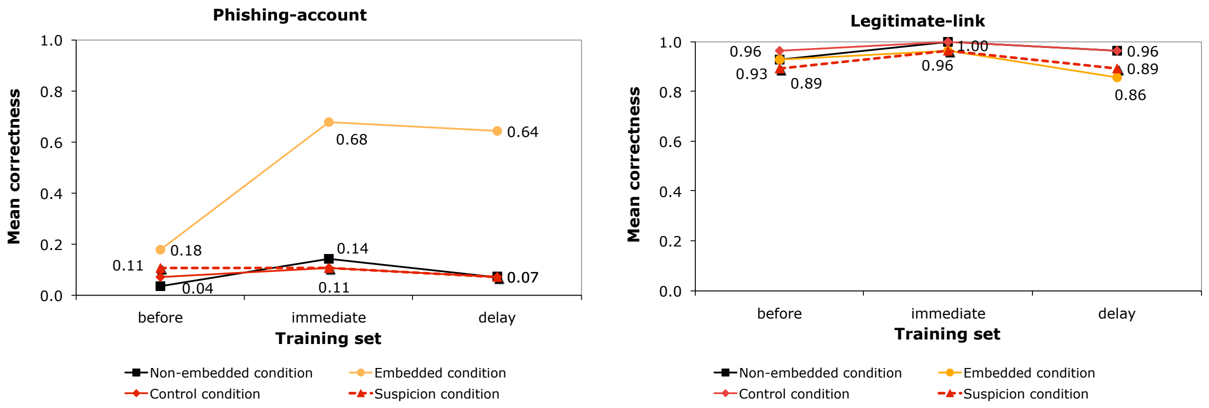


Figure 6.7: Mean correctness for identifying phishing-account (left) and legitimate-link (right) emails before training, immediately after training, and after a one-week delay. Left figure shows that participants in the embedded condition did significantly better immediately and after a delay than participants in the other conditions. Right figure shows that training does not increase the false positive error.

Participant scores and behavior

For each participant, we calculated a score between 0 and 7 on each email set. To determine the score, we counted the number of correct decisions that participants made about both the spam email and the set of emails that contained links. We counted a decision about a legitimate email as correct if the participant clicked on the link and performed the requested action. We counted a decision about a phishing email as correct if the participant did not click on the link in that email. We counted a decision about a spam email as correct if the participant did not open the email. We also calculated the percentage correct for each participant and each type of email in each set.

Before the training, we found no significant difference (t = 1.48, p-value = 0.17) in scores for the phishing-account messages in email sets A and C, indicating that they were of similar difficulty. Within each group, we found no significant difference between the scores for the two phishing-account emails that the participants received (proportion test: A group, p-value = 0.37, and C group, p-value = 0.32). This shows that the phishing-account emails presented in each group did not differ significantly.

Among the seven participants excluded from the analysis because they did not look at the training materials, three were in the non-embedded condition and four were in the embedded condition. Among the four in the embedded condition, two participants did not open the training email and two of them did not click on the link in the email. None of the three participants in the non-embedded condition who did not look at training materials opened the email. The total correctness score for participants who did not look at the intervention was 6.33 for the embedded condition and 6.25 for the non-embedded condition. We found a significant difference between the scores of people who saw the training and people who did not see the training material. The responses of these seven participants are not included in the analysis discussed in this section.

We found no significant correlation between phishing susceptibility and the demographic information we collected. For instance, there was no significant correlation between participants' age and total scores (Pearson coefficient r = 0.30, p-value = 0.13). There was no significant correlation between emails received per week (excluding unsolicited) and total scores (Pearson coefficient r = 0.02, p-value = 0.92). There was no significant correlation between shopping online in the last six months and total scores (Pearson coefficient r = -0.12, p-value = 0.56). There was no significant correlation between hours of Internet usage per week and total scores (Pearson coefficient r = 0.24, p-value = 0.22). There was also no significant difference in scores between males and females (t = -1.1, p-value = 0.29). The mean score for males was 4.27 (s.d = 1.19, var = 1.42) and the mean score for females was 4.71 (s.d = 0.69, var = 0.47). We also observed no significant difference between the non-embedded condition and the control condition (details in Figure 6.7).

Learning

In this section we assess how much users learned as a result of the interventions.

User performance: To test hypothesis 1, we evaluated the effectiveness of the training by looking at the percentage of correct responses for each participant. This evaluation was performed in each condition for phishing and legitimate-link emails both before and after the training.

Participants in the embedded and non-embedded conditions did not perform significantly differently when it came to correctly identifying phishing-account emails before the training (two sample t-test: df = 26, p-value = 0.19). However, those in the embedded condition performed significantly better than those in the non-embedded condition immediately after training (two sample t-test: df = 26, p-value < 0.01), as shown in Figure 6.7. At that time, those in the embedded condition improved their performance significantly (paired t-test: t = -3.61, df = 13, p-value < 0.01), while those in the non-embedded condition did not (paired t-test: t = -1.15, df = 13, p-value = 0.27). There was no significant difference between the control condition and the non-embedded condition both before and after the training. There was also no significant difference between the suspicion condition and the non-embedded condition both before and after the training.

Participants in the embedded and non-embedded conditions did not perform significantly differently when it came to correctly identifying legitimate-link emails before or after the training, as shown in Figure 6.7. There was no significant difference in mean correctness before and immediately after the training for embedded (paired t-test: t = -1, df = 13, p-value = 0.34) and non-embedded conditions (paired t-test: t = -1.47, df = 13, p-value = 0.17). Similarly, there was no significant difference for mean correctness between the non-embedded and control conditions or between the non-embedded and suspicion conditions.

These results supported Hypothesis 1, demonstrating that embedded training increases users' ability to detect phishing-account emails while non-embedded training does not. No form of training had significant impact on users' ability to recognize legitimate emails.

Time spent in reading the intervention: One approximate measure for how closely people read the training materials is the time they spend looking at the materials. Learning science suggests that users exposed to training materials for more time may learn more [198]. We measured the time participants spent on the training materials in each condition. There was significant difference (two sample t-test: t = -3, df = 26, p-value < 0.01) between the embedded condition (min = 21 seconds, max = 240 seconds, avg = 97 seconds) and the non-embedded condition (min = 2 seconds, max = 100 seconds, avg. = 37 seconds). This shows that participants in the embedded condition spent significantly more time reading the training material than participants in the non-embedded condition. We also found significant correlation between time spent reading the training material and both the total scores immediately after the training (Pearson coefficient r = 0.6, p-value < 0.01) and the scores after the delay (Pearson coefficient r = 0.44, p-value = 0.02).

Retention and transfer

In order to measure retention and transfer, we asked participants to come back for a second part of the study. We requested that they come back exactly 7 days after part 1. However, not all of the participants came back in exactly seven days. The participants from the non-embedded condition came an average of 7.5 days apart (min = 6, max = 9, s.d = 0.94, var = 0.88). Embedded condition participants on average came back after 7.2 days (min = 6, max = 9, s.d = 0.80, var = 0.64). Control condition participants on average came back after 7.1 days (min = 6, max = 9, s.d = 0.7, var = 0.5). Suspicion condition participants on average came back after 7.1 days (min = 6, max = 8, s.d = 0.7, var = 0.5). There was no significant difference in days apart between the four conditions.

Overall performance after a delay: In order to measure overall user performance after the one-week delay, we compared correctness percentages for phishing-account and legitimate-link emails before training, immediately after training, and after a one-week delay. As shown in Figure 6.7, participants in the embedded condition performed significantly better then those in the non-embedded condition even after the one-week delay (two sample t-test: df = 26, p-value < 0.01). Participants in the embedded condition performed significantly better after the delay than they had before training (paired t-test: t = -2.51, df = 13, p-value = 0.02); participants in the non-embedded group failed to improve their performance (paired t-test: t = -0.43, df = 13, p-value = 0.67). In both conditions there was no significant difference between performances immediately after the training and after a delay of one-week. Participants in the control and suspicion condition did not perform significantly better after the delay than they had immediately after training.

As shown in Figure 6.7, participants in the embedded, non-embedded, suspicion and control conditions did not perform significantly differently in correctly identifying legitimate-link emails after the delay. There was no significant difference in mean correctness between performances before the training and performances after the delay in all four conditions.

These results suggest that users were able to correctly identify phishing and legitimate emails better in the embedded condition than in the non-embedded, suspicion and control conditions even after a one-week delay.

Retention: During the study, an email was sent that appeared to be from Amazon, with the subject "Revision to your Amazon.com information." This email asked the user to update the personal information on their Amazon account. To measure retention, we used an email from Citibank that was similar to the Amazon email; this email asked users to update their personal information for the account. There was a significant difference between participants from the non-embedded and embedded training conditions when it came to correctly identifying the email from Citibank as phishing email (two sample t-test: df = 26, p-value < 0.01). There was also significant difference between the embedded condition and the control and suspicion conditions. This result lent support to Hypothesis 2. Only 7% of participants identified the email correctly in the non-embedded, suspicion and control conditions, while 64% of participants identified the email correctly in the embedded condition. One of the participants in the embedded condition mentioned that "I remember reading last time that thing [training material] said not click and give personal information."

Transfer: To measure the knowledge transfer, we used an phishing-account type email that asked participants to reactivate their eBay account. This email type is different from the Amazon account email that they received. We found significant differences between the non-embedded and the embedded training conditions in terms of correctly identifying the eBay email as a phishing attack (two sample t-test: df = 26, p-value < 0.01). This result lend support to Hypothesis 3. Only 7% of the participants identified the email correctly in the non-embedded, suspicion and control conditions, while 64% of the participants identified the email correctly in the embedded condition. One of the participants in the embedded condition mentioned that "PhishGuru said not to click on links and give personal information, so I will not do it, I will delete this email."

Cognitive Reflection

As mentioned earlier, we included Frederick's three-question Cognitive Reflection Test (CRT) in the pre-screening survey. The raw CRT score ranged from 0 to 3, with "0" indicating that the subject did not answer any of the three questions correctly and "3" indicating that the subject answered all three correctly. The mean CRT score was 1.2 and s.d. = 0.9. We dichotomized the CRT score by converting CRT scores of 0−1 to "low CRT group" and 2−3 to "high CRT group." We had 33 subjects in the low CRT group and 23 in the high CRT group. There was no significant difference between the means of each of the four conditions. We also found no significant correlation between the age of the participants and the CRT score (Pearson coefficient r $= -0.2$, p-value $= 0.4$).

We tested the cognitive reflection hypotheses by comparing the proportion of individuals in the two CRT groups (high and low) who clicked on the phishing-account and phishing-no-account emails prior to training; to do this, we used a test of two proportions. For Hypothesis 4, we predicted that the high CRT group had a lower probability of clicking on the phishing email sent by from the company with whom they had an account. Using a test of 2-proportions, we found a difference in the predicted direction; however, the statistical analysis suggested that this difference between the proportion of the low CRT group (0.85) and the high CRT group (0.68) who clicked on the phishing email is not significant (proportion test: p-value $= 0.14$). This result rejects Hypothesis 4.

In the case of Hypothesis 5, we expected that subjects who had higher CRT scores would be more likely to click on the phishing-no-account emails prior to training. We made this conjecture because if higher CRT scores correlate with more risk-taking, then high CRT subjects should be more likely to click on unexpected emails (given the Bobby Smith storyline). In the sample, the high CRT group had a higher probability of clicking on the phishing-no-account e-mails than those in the high CRT group, 0.43 versus 0.04, respectively. A test of 2-proportions suggests that the difference in proportions was significant (p-value < 0.01). These results indicate that those with high CRT scores are more likely to click on phishing-no-account e-mails than those with low CRT scores. This result lends support to Hypothesis 5. It does not mean that those who are more "cognitively reflective" are more likely to fall for phishing attacks, but may suggest that they are more inclined to "play with fire." In a novel situation, they may be more inclined to experiment with unknown e-mails than those with lower CRT scores. However, this may or may not suggest that those with high CRT scores are more likely to be "burned." In a real situation, although they may be curious about the e-mail, its content, and the website it links to, they may not necessarily enter their personal information on a website they do not trust. Nevertheless, clicking on the email may expose individuals to other types of security threats such as viruses.

Observations

Participants in all conditions identified spam emails correctly most of the time, not even opening them. Almost all (93%) of the participants identified the spam email correctly before training in all conditions. One of the participants who opened the spam email was curious about it (subject of the email: "Fw: Re: You will want this Job"). Another participant said "Oh, it is offering me a job, might be interesting, let me see it." There was no significant difference within the conditions for before training, immediately after training, and after a delay. There was also no significant difference between the conditions in any of the states.

Among participants, there was a significant difference in correctness for the phishing emails from organizations they had an account with (phishing-account) and those with whom they did not have an account (phishing-no-account). There was a significant difference between the phishing-account and phishing-no-account emails within the conditions before the training. One of the common reasons mentioned by the participants for not opening or for deleting the phishing-no-account emails is "I don't have an account with this organization." In particular, one of the participants mentioned, "I don't have account with Barclays, how did they get my email address, and why are they sending emails asking me to update my information?"

We observed that participants in the embedded condition were motivated to read the training material longer than those in the non-embedded condition. One participant mentioned, "I was more motivated to read the training materials since it was presented after me falling for the attack." This quote succinctly captures the motivation behind the embedded training methodology, which makes training part of users' primary task. Another participant in the embedded condition mentioned, "Thank you PhishGuru, I will remember that [the 5 instructions given in the training material]." In general, participants who spent time reading the training material liked the design. One participant who was not aware that URLs could be misleading looked at the arrow pointing to the first n in anazon.com (Figure 6.6) and said, "That is scary, I will be careful in the future. That [instruction] is good to know." The non-embedded condition however, did not inspire the same motivation as the embedded condition. As one of the participants commented "This [image in the email] looks like some spam." Another participant mentioned "I dont know why Amazon would send me such [intervention] in the email."

6.2.5 Discussion

In this section we showed that: (1) users learned more effectively when the training materials were presented after they fell for the phishing attack (embedded) than when the training materials were sent by email (non-embedded); (2) users retained more knowledge when they were trained with embedded training than when they were trained with non-embedded training; (3) users transferred more knowledge about how to avoid phishing attacks when they were trained with embedded training than when they were trained with non-embedded training; (4) users with high and low Cognitive Reflection Test (CRT) scores had an equal likelihood of clicking on links in the phishing emails from organizations they had an account with (phishing-account emails); and (5) users with high CRT scores were more likely than users with low scores to click on links in emails from organizations that they did not have an account with (phishing-no-account emails); these users may have been motivated by curiosity.

The results from the study supported hypotheses 1, 2, 3 and 5, and rejected hypothesis 4. Results from this study contradicted the conventional wisdom that it is hard to train novice users about security. The results are consistent with the findings of learning science, which suggest that users can be trained if the training methodology is systematically designed and applies learning science principles.

These results strongly suggest that sending instructional materials through email (non-embedded) does not motivate users to spend time reading them. We believe this is because people do not understand why they are receiving such emails and so delete the emails with the instructions. The results also suggest that users are motivated to learn when training materials are presented after users fall for phishing emails (that is, when users click on the link in the email). We believe this is because the embedded methodology directly applies the principles of learning-by-doing and immediate feedback.

These results suggest that users can retain and transfer knowledge if they are motivated to read training materials. After seven days, participants in the embedded condition retained learned knowledge better than participants in the non-embedded condition. This may suggest that inspiring motivation by making users fall for phishing emails influences their retention of knowledge. We also found that participants in the embedded condition were able to transfer their knowledge to a situation different from the training situation better than participants in the non-embedded condition. This suggests that when users receive frequent training on phishing emails, they should be able to identify other types of phishing emails.

In the post-study discussion with participants, almost all participants liked the comic script intervention design we used for this study. We attribute this to the learning science principles (learning-by-doing, immediate feedback, contiguity, personalization, and story-based agent) we applied when creating the design.

According to our analysis, users with high and low CRT scores were equally likely to click on links in the phishing emails from organizations they had an account with. This analysis also found that participants with high CRT scores were more likely to click on phishing emails from an unknown source. This result indicates that it may be appropriate to train the high CRT score group to not click on links from unknown sources.

In this chapter, using laboratory studies, we showed that PhishGuru effectively trains users to identify phishing emails immediately after being trained and after 7 days. We also showed that users retained more knowledge when they were trained with PhishGuru than when they were trained with non-embedded training. Using the lessons learned from these two laboratory studies, we conducted two real-world studies, which are discussed in the next chapter.

Chapter 7

Real World Evaluation of PhishGuru

As discussed in Chapter 6, prior laboratory studies showed that PhishGuru, an embedded training system, is an effective way to teach users to identify phishing scams. However, laboratory studies are unable to fully replicate real world conditions: they may lack ecological validity and fail to sufficiently approximate real-world situations. This in turn may impact external validity – that is, the ability to make generalized inferences from the results [29].

Laboratory studies are very helpful to researchers who wish to understand user behavior in a given situation. However, most laboratory studies have tradeoffs and face validity challenges: they contend with both ecological (whether the methods, materials, and settings are similar to real life) and external (whether the results are generalizable) validity issues [29]. Laboratory studies in the context of phishing also grapple with ethical issues: how much the researcher should inform the participant about the study and how much deception is acceptable [95, 100]. In one laboratory experimental setup, researchers showed that people who role-play behave differently than people who use their own credentials [173].

Few real world studies of users' behavior in the context of phishing have been conducted; even fewer real world studies have been conducted to evaluate the effectiveness of anti-phishing training. Real world evaluations of anti-phishing training that have been conducted involved classroom and office training as well as training delivered via an online game [181]. Real world studies have been used to evaluate participants' susceptibility to phishing, but not to evaluate the effectiveness of training [64, 90, 151]. To develop effective countermeasures for phishing, researchers must understand users' behavior in real world settings. Even though real world studies provide richer data, it can be difficult to control a real-world study setup due to the many sources of variability [168]. It can also be difficult to make arrangements for a real-world study, especially when it requires a company to cooperate by providing access to employees or customers. Companies may not grant desired access or permit publication of study data or results. Real world studies also pose ethical challenges, as they must often be conducted without obtaining prior consent from individual participants [95,100].

The focus of this chapter is to build on earlier PhishGuru laboratory studies by conducting two studies in a real world setting. In Section 7.1, we discuss a study conducted with employees of a Portuguese company. In Section 7.2, we discuss a study administered among the staff, faculty, and students of Carnegie Mellon University.

7.1 First evaluation: Portuguese company

> This section is largely a reproduction of a paper co-authored with Steve Sheng, Alessandro Acquisti, Lorrie Cranor, and Jason Hong and published at e-Crime Researchers Summit 2008 [113].

In this section, we focus on the study conducted with a Portuguese company. The remainder of this section is organized as follows: In Section 7.1.1, we describe the study setup and participant demographics. In Section 7.1.2, we present the hypotheses that guided this study. In Section 7.1.3, we present the results of the evaluation, demonstrating that PhishGuru effectively educates people in the real world. In Section 7.1.4, we discuss the effect of training people in the real world.

7.1.1 Study setup and participant demographics

This study was conducted at a large Portuguese company. All emails and training materials were translated into Portuguese. All participants in the study worked on the same floor of an office building, but came from different departments: administration, business, design, editorial, management, technical, and others.

The study included three conditions: "control," "generic training," and "spear training." Participants in the control condition did not receive any training. Participants in the generic training condition received a simulated spear phishing email (targeted phishing email) and saw generic phish training material (Figure 7.1) when they clicked on a link in the email. Participants in the spear training condition received a simulated spear phishing email and saw spear phish training material (Figure 7.2) when they clicked on a link in the email. We assigned 111 employees to the control condition, 100 to the generic training condition, and 100 to the spear training condition. Table 7.1 presents the demographics of the study participants.

The company we worked with was primarily interested in studying the vulnerability of their employees to spear phishing emails, so we used spear phishing emails for all simulated phishing emails in this study. Targeted spear phishing attacks have been more successful than generic phishing attacks at conning people and causing damage to companies and individuals.

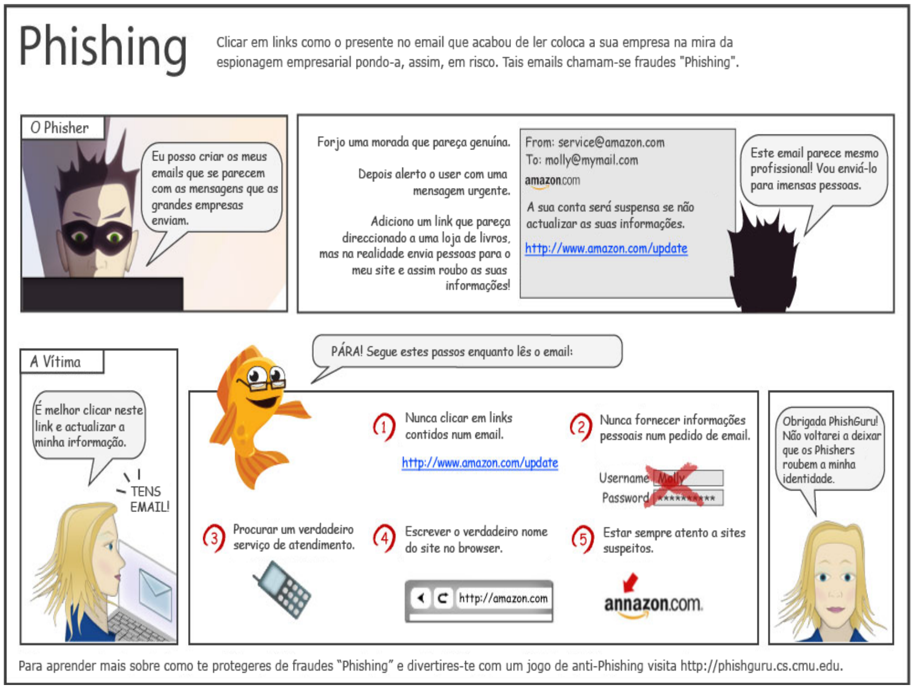

Phishing
Clicar em links como o presente no email que acabou de ler coloca a sua empresa na mira da espionagem empresarial pondo-a, assim, em risco. Tais emails chamam-se fraudes "Phishing".
O Phisher
Eu posso criar os meus emails que se parecem com as mensagens que as grandes empresas enviam.
Forjo uma morada que pareça genuína.
Depois alerto o user com uma mensagem urgente.
Adiciono um link que pareça direccionado a uma loja de livros, mas na realidade envia pessoas para o meu site e assim roubo as suas informações!
From: service@amazon.com
To: molly@mymail.com
amazon.com
A sua conta será suspensa se não actualizar as suas informações.
http://www.amazon.com/update
Este email parece mesmo profissional! Vou enviá-lo para imensas pessoas.
A Vítima
É melhor clicar neste link e actualizar a minha informação.
TENS EMAIL!
PÁRA! Segue estes passos enquanto lês o email:
1 Nunca clicar em links contidos num email.
http://www.amazon.com/update
2 Nunca fornecer informações pessoais num pedido de email.
Username Molly
Password **********
3 Procurar um verdadeiro serviço de atendimento.
4 Escrever o verdadeiro nome do site no browser.
http://amazon.com
5 Estar sempre atento a sites suspeitos.
annazon.com
Obrigada PhishGuru! Não voltarei a deixar que os Phishers roubem a minha identidade.
Para aprender mais sobre como te protegeres de fraudes "Phishing" e divertires-te com um jogo de anti-Phishing visita http://phishguru.cs.cmu.edu.

Figure 7.1: Generic intervention in Portuguese. English version presented in Figure 7.3.

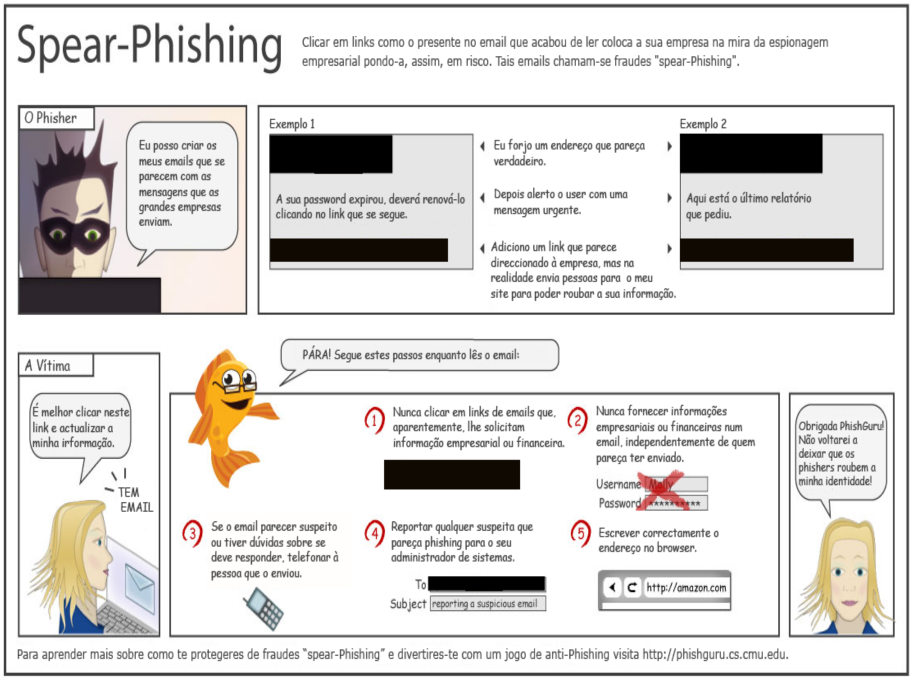

Spear-Phishing
Clicar em links como o presente no email que acabou de ler coloca a sua empresa na mira da espionagem empresarial pondo-a, assim, em risco. Tais emails chamam-se fraudes "spear-Phishing".
O Phisher
Eu posso criar os meus emails que se parecem com as mensagens que as grandes empresas enviam.
Exemplo 1
[redacted]
A sua password expirou, deverá renová-lo clicando no link que se segue.
[redacted]
Eu forjo um endereço que pareça verdadeiro.
Depois alerto o user com uma mensagem urgente.
Adiciono um link que parece direccionado à empresa, mas na realidade envia pessoas para o meu site para poder roubar a sua informação.
Exemplo 2
[redacted]
Aqui está o último relatório que pediu.
[redacted]
A Vítima
É melhor clicar neste link e actualizar a minha irformação.
TEM EMAIL
PÁRA! Segue estes passos enquanto lês o email:
1 Nunca clicar em links de emails que, aparentemente, lhe solicitam informação empresarial ou financeira.
[redacted]
2 Nunca fornecer informações empresariais ou financeiras num email, independentemente de quem pareça ter enviado.
Username Molly
Password **********
3 Se o email parecer suspeito ou tiver dúvidas sobre se deve responder, telefonar à pessoa que o enviou.
4 Reportar qualquer suspeita que pareça phishing para o seu administrador de sistemas.
To [redacted]
Subject reporting a suspicious email
5 Escrever correctamente o endereço no browser.
http://amazon.com
Obrigada PhishGuru! Não voltarei a deixar que os phishers roubem a minha identidade!
Para aprender mais sobre como te protegeres de fraudes "spear-Phishing" e divertires-te com um jogo de anti-Phishing visita http://phishguru.cs.cmu.edu.

Figure 7.2: Spear intervention in Portuguese. English version presented in Figure 7.4.

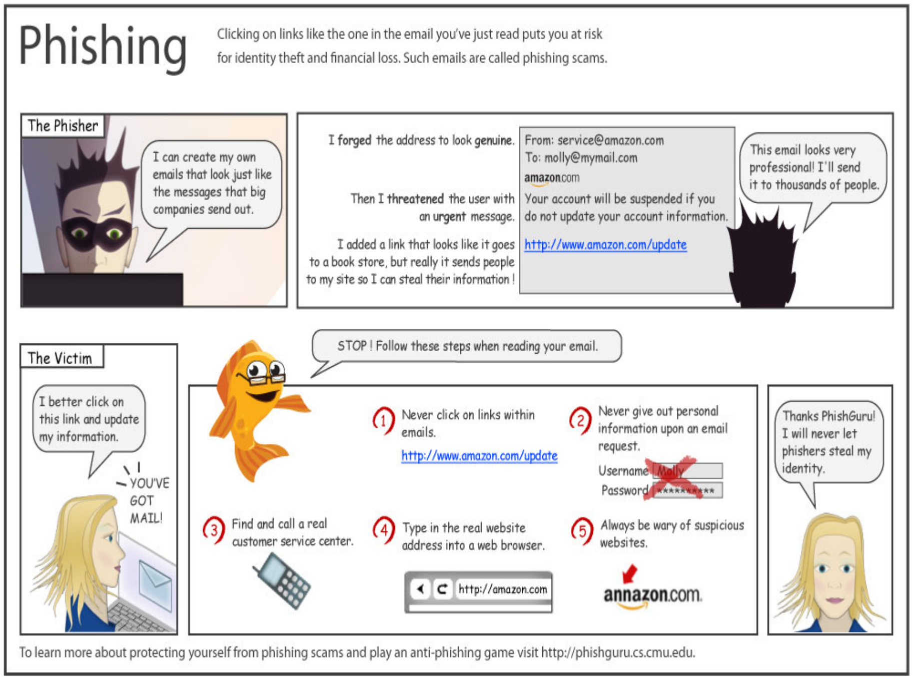

Phishing
Clicking on links like the one in the email you've just read puts you at risk for identity theft and financial loss. Such emails are called phishing scams.
The Phisher
I can create my own emails that look just like the messages that big companies send out.
I forged the address to look genuine.
From: service@amazon.com
To: molly@mymail.com
amazon.com
Then I threatened the user with an urgent message.
Your account will be suspended if you do not update your account information.
I added a link that looks like it goes to a book store, but really it sends people to my site so I can steal their information!
http://www.amazon.com/update
This email looks very professional! I'll send it to thousands of people.
The Victim
I better click on this link and update my information.
YOU'VE GOT MAIL!
STOP! Follow these steps when reading your email.
1 Never click on links within emails.
http://www.amazon.com/update
2 Never give out personal information upon an email request.
Username Molly
Password **********
3 Find and call a real customer service center.
4 Type in the real website address into a web browser.
http://amazon.com
5 Always be wary of suspicious websites.
amazon.com
Thanks PhishGuru! I will never let phishers steal my identity.
To learn more about protecting yourself from phishing scams and play an anti-phishing game visit http://phishguru.cs.cmu.edu.

Figure 7.3: Generic intervention. English version of Figure 7.1.

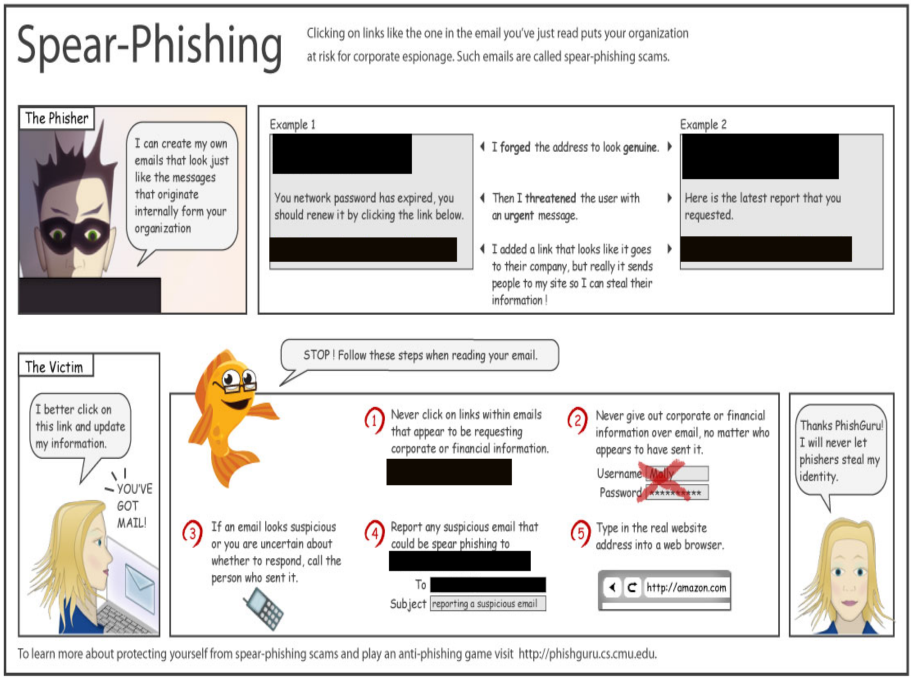

Spear-Phishing
Clicking on links like the one in the email you've just read puts your organization at risk for corporate espionage. Such emails are called spear-phishing scams.
The Phisher
I can create my own emails that look just like the messages that originate internally form your organization
Example 1
[redacted]
You network password has expired, you should renew it by clicking the link below.
[redacted]
I forged the address to look genuine.
Then I threatened the user with an urgent message.
I added a link that looks like it goes to their company, but really it sends people to my site so I can steal their information!
Example 2
[redacted]
Here is the latest report that you requested.
[redacted]
The Victim
I better click on this link and update my information.
YOU'VE GOT MAIL!
STOP! Follow these steps when reading your email.
1 Never click on links within emails that appear to be requesting corporate or financial information.
[redacted]
2 Never give out corporate or financial information over email, no matter who appears to have sent it.
Username Molly
Password **********
3 If an email looks suspicious or you are uncertain about whether to respond, call the person who sent it.
4 Report any suspicious email that could be spear phishing to
[redacted]
To [redacted]
Subject reporting a suspicious email
5 Type in the real website address into a web browser.
http://amazon.com
Thanks PhishGuru! I will never let phishers steal my identity.
To learn more about protecting yourself from spear-phishing scams and play an anti-phishing game visit http://phishguru.cs.cmu.edu.

Figure 7.4: Spear intervention. English version of Figure 7.2.

Table 7.1: Participant demographics.

	Control Condition (N=111)	Generic training condition (N=100)	Spear training condition (N=100)
Gender			
Male	77%	27%	67%
Female	23%	73%	33%
Areas of work			
Administration	1%	1%	1%
Business	2.7%	5%	9%
Design	5.4%	3%	7%
Editorial	4.5%	5%	7%
Management	22.5%	19%	20%
Technical	39.6%	36%	35%
Others	24.3%	31%	21%

In total, participants received four emails during the study: three simulated spear phishing emails and one legitimate email containing a link. The spear phishing emails and the legitimate email were all based on actual emails that the company had received or the kind of emails that the system administrators were worried about.

The first email employees received was a training email (Train) delivered on day 0 to employees in the generic and spear conditions. It was a spear phishing email that asked employees to log into the corporate network by clicking on a link and entering their user name and password. When employees clicked on the link in this email, they were taken to the training material corresponding to the condition they were in. Participants in the generic training condition saw the generic phish training message shown in Figure 7.1, while participants in the spear training condition saw the spear phish training message shown in Figure 7.2.

The second email (Test 1) was designed to measure the knowledge employees acquired through our training materials. In order to compare trained and untrained employees, this email was sent to employees in all conditions. To measure immediate effectiveness, this email was sent on day 2 of the study. This simulated spear phishing email said that the recipient's internal network password had expired and asked them to click on a link and change their password. When employees clicked on the link in this email, they were taken to a fake phishing website that looked the same as the real website and was hosted on a similar-looking domain name.

The third email (Test 2), which was designed to measure retention, was sent on day 7. As in Test 1, this email was sent to participants in all conditions to compare the trained and untrained employees. This email asked employees to click on a link and update their communication information for internal corporate communication purposes. When employees clicked on the link they were taken to a phishing website that looked the same as the real website and was hosted on a similar looking domain name.

We also wanted to find out if training increases participants' concern level to such a high degree that they stop clicking on any links, even legitimate ones. Testing this possibility, we sent a legitimate email with a link (Test 3) on day 10 to all participants in all conditions. This email asked employees to click on a link to read the company's updated security policy. When employees clicked on the link, they were taken to a legitimate webpage with the updated security policy. Table 7.2 summarizes all emails, email types, days on which the email was sent, conditions under which the emails were delivered, and relevant features of the email.

The phishing websites that participants saw when they followed the links in the spear phishing emails were exact replicas of real company websites. However, they were hosted on a domain that looked similar to but not the same as the company's domain. All replicated websites were completely functional and allowed employees to submit information. So that only company employees could access the training materials and fake phishing websites, these websites were hosted in a way that only granted access to IP addresses coming from the company's domain. This also helped us identify the IP address and thereby the user from whose machine the request had come. The company tracked all of this information; for privacy reasons, we did not receive specific details like the IP address, etc. from the company. We tracked the clicks to the phishing websites and the training materials, as well as the information submitted to the phishing websites.

To make sure the employees received the emails that were part of the study, system administrators bypassed the corporate email filters and placed them in participants' inboxes.

All participants were asked to complete a post-study survey on day 20. The survey consisted of questions regarding (1) the interest level of participants in receiving such emails in future; (2) participants' feedback on the training; and (3) participants' feedback on the instructions.

7.1.2 Hypotheses

In this section we introduce three hypotheses which informed the study.

Replicating laboratory study results

Earlier laboratory studies have shown that a large percentage of participants who click on links in simulated emails proceed to give some form of personal information to the phishing website. As seen in Chapter 6, this percentage was around 90% in laboratory studies. The goal here was to investigate whether this holds true in a real world setting. This result may show that people have to be trained to not click on links; otherwise, there is a low probability that they will click on links

Table 7.2: Summary of emails sent to study participants.

Emails	Type	Day of sending	Conditions	Relevant features of the email
Train	Spear phishing	Day 0	Generic and spear	Asked user to enter their user name and password in order to use the corporate network
Test 1	Spear phishing	Day 2	All	Told user their internal network password had expired; asked them to change their password
Test 2	Spear phishing	Day 7	All	Asked user to update their communication information
Test 3	Legitimate with link	Day 10	All	Asked user to read the company's updated security policy

and not go on to give personal information to phishing websites.

> **Hypothesis 1**: In the real world, a large percentage of people who click on links in simulated emails go on to provide some form of personal information.

An earlier laboratory study (discussed in Section 6.2) showed that users learn, retain, and transfer effectively when training materials are presented after they fall for a phishing attack. The goal here was to investigate whether this was true in a real world setting.

> **Hypothesis 2**: PhishGuru (embedded training) effectively trains people in the real world.

To evaluate the effectiveness of PhishGuru, we calculated the following: (1) percentage of participants who clicked on a link in phishing emails and gave information to fake phishing websites immediately after training; (2) percentage of participants who clicked on a link in phishing emails and gave information to fake phishing websites 7 days after training; and (3) percentage of participants who clicked on a link in legitimate emails after training.

Generic and spear training instructions

The content of training materials affects the way people learn and reproduce knowledge. Researchers have shown that people make better decisions if the testing situation is the same or similar to the training situation and training materials than if the testing situation is different [44].

To investigate the effect of this difference in the instructions, we developed one set of anti-phishing instructions that were generic and another specific to spear phishing emails. Figure 7.1 and Figure 7.2 have the same content except for the instructions in the lower pane of the material. As the training materials used in the study were in Portuguese, the translated English version of the instructions is given in Table 7.3. The English version of the messages is given in Figure 7.3 and Figure 7.4.

> **Hypothesis 3**: People trained with spear training material can better identify spear phishing emails than people trained with generic training material.

Table 7.3: Translated English version of the instructions in the training materials.

Generic training instructions	Spear training instructions
1.Never click on links within emails.	1.Never click on links within emails that appear to be requesting corporate or financial information.
2.Never give out personal information upon email request.	2.Never give your corporate or financial information over the email, no matter who appears to have sent it.
3.Find and call a real customer service center.	3.If an email looks suspicious or you are uncertain about whether to respond, call the person who sent it.
4.Type in the real website address into a web browser.	4.Report any suspicious email that could be spear phishing to sysadmin@company.com.
5.Always be wary of suspicious websites.	5.Type in the real website address into a web browser.

7.1.3 Results

In this section we present the results of the study. The results from this study supported Hypotheses 1 and 2, but not Hypothesis 3. We found that a large percentage of the participants who clicked on links in simulated emails gave away some form of personal information to the fake phishing websites that were part of the study. We found that participants in the training conditions made significantly better decisions after the training than they did before the training. Results from this study suggest that users retained knowledge gained from PhishGuru for at least 7 days after the training. However, the difference in the instructions in the training materials did not have a significant effect on the participants' ability to identify phishing emails. Surprisingly, the results also suggest that PhishGuru training could effectively train other people in the organization who did not receive training messages directly from the system. The complete decision tree for all three conditions is given in Figure 7.5, Figure 7.6, and Figure 7.7.

Table 7.4: Percentage of participants who clicked on the training link, only clicked, and clicked and gave information on other emails.

Conditions	Clicked on link in training email on day 0	Clicked on link on day 2	Clicked on link and gave information on day 2	Clicked on link on day 7	Clicked on link and gave information on day 7
Control	N/A	20 %	19 %	17 %	15 %
Generic training	42 %	17 %	15 %	14 %	12 %
Spear training	39 %	14 %	12 %	17 %	14 %

Giving away personal information

In this study, we found that a large percentage of participants who clicked on links in simulated phishing emails went on to give some form of personal information to the phishing websites. The system administrators in the company who helped us conduct the study had access to the information that was entered into phishing websites. They were able to check the usernames and other details that were entered. We found that 88% of the participants who clicked on links went on to give some form of personal information to the fake phishing websites. In the earlier laboratory studies, we found that 90 to 93 percent of participants who clicked on links gave their personal information to fake phishing websites (Chapter 6). Table 7.4 gives the percentage of participants in each condition who clicked on a link in phishing emails; it also lists the percentage who clicked and gave information to fake phishing websites.

Phishing emails

We found that PhishGuru training led participants to make better decisions relating to phishing emails they received as part of the study. Before training (see Table 7.4), there was no significant difference between the generic (42%) and spear (39%) training conditions in the percentage of participants who clicked on the link in the phishing email and gave information (two sample T-test, p-value = 0.6). This shows that, before the training, participants were at the same level in both conditions.

In both training conditions (generic and spear), participants made better decisions immediately after training. We found that (see Table 7.4), in the generic condition, the percentage of participants who clicked and gave information dropped significantly, from 42% on day 0 to 15% on day 2 (paired

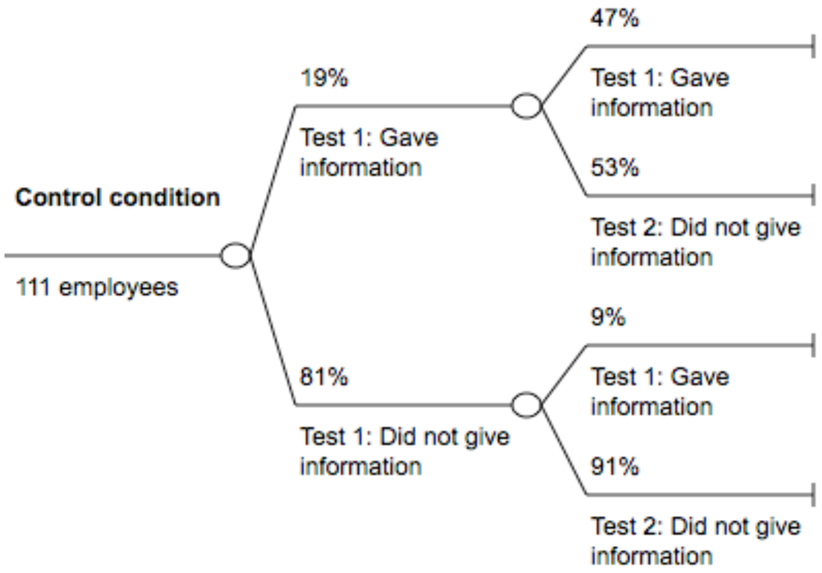


Figure 7.5: Decision tree for control condition. It presents the percentage of employees who clicked on links in the email and gave information and the percentage of employees who did not click on links.

T-test, $p\text{-value} < 0.01$). In the spear training condition, the percentage also decreased significantly, from 39% on day 0 to 12% on day 2 (paired T-test, $p\text{-value} < 0.01$).

Trained participants (who clicked on the link in the Train email and saw the training materials) retained the knowledge gained from PhishGuru training for at least 7 days. Table 7.5 lists the percentage of trained participants who clicked on the link and gave information. The untrained group includes participants from both the generic training and spear training conditions who did not see the training materials. As Table 7.5 shows, participants in the generic training (Paired T-test, $p\text{-value} = 0.55$) and spear training (Paired T-test, $p\text{-value} = 0.67$) conditions did not perform significantly worse on day 7 than they did on day 2.

We found that a significant number of trained participants correctly identified both test emails. Table 7.6 shows the percentage of control, trained, and untrained participants who identified day 2 and day 7 emails correctly. The untrained group includes participants from both the generic and spear training conditions who did not see the training materials because they did not click on the link in the first phishing email. In the trained conditions, a significant number of participants identified both emails correctly. We believe that additional training with a second training email could further improve the percentage of participants able to correctly identify both emails. Results also showed that untrained participants identified phishing emails better than trained participants. This suggests that most of these untrained participants did not need the training they did not receive.

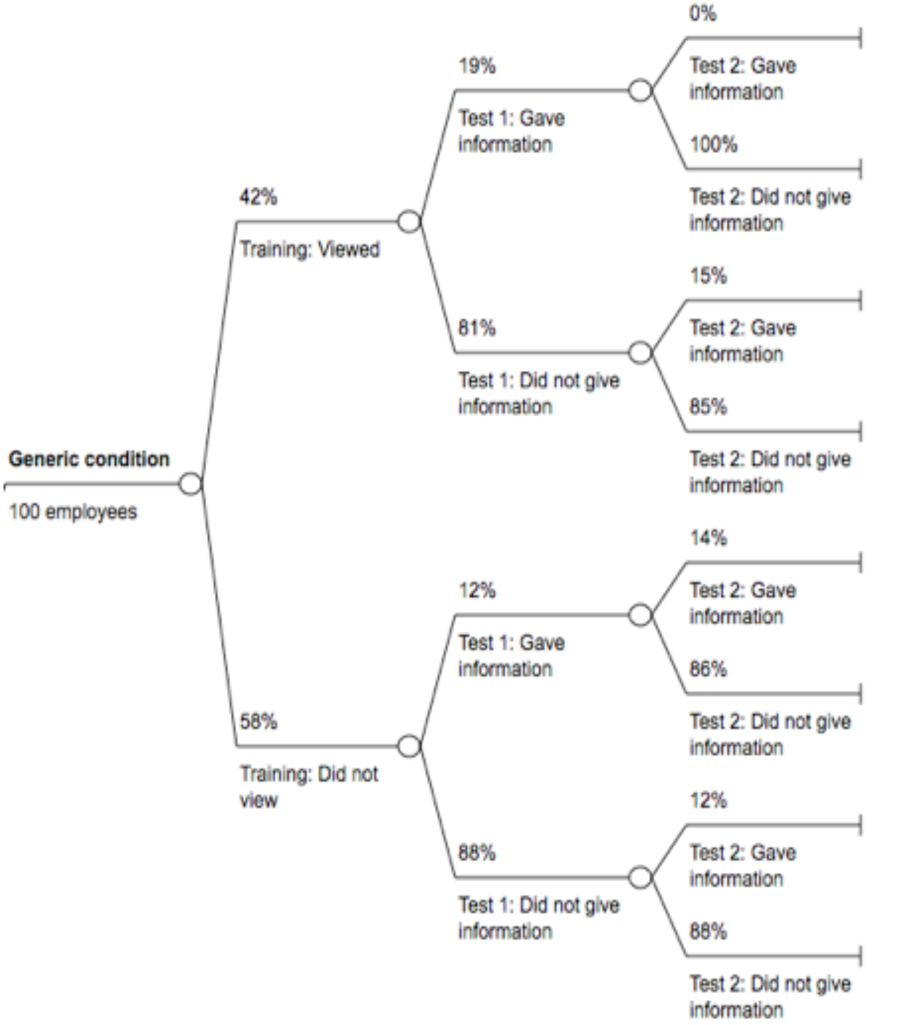


Figure 7.6: Decision tree for generic condition. It presents the percentage of employees who clicked on links in the email and gave information and the percentage of employees who did not click on links.

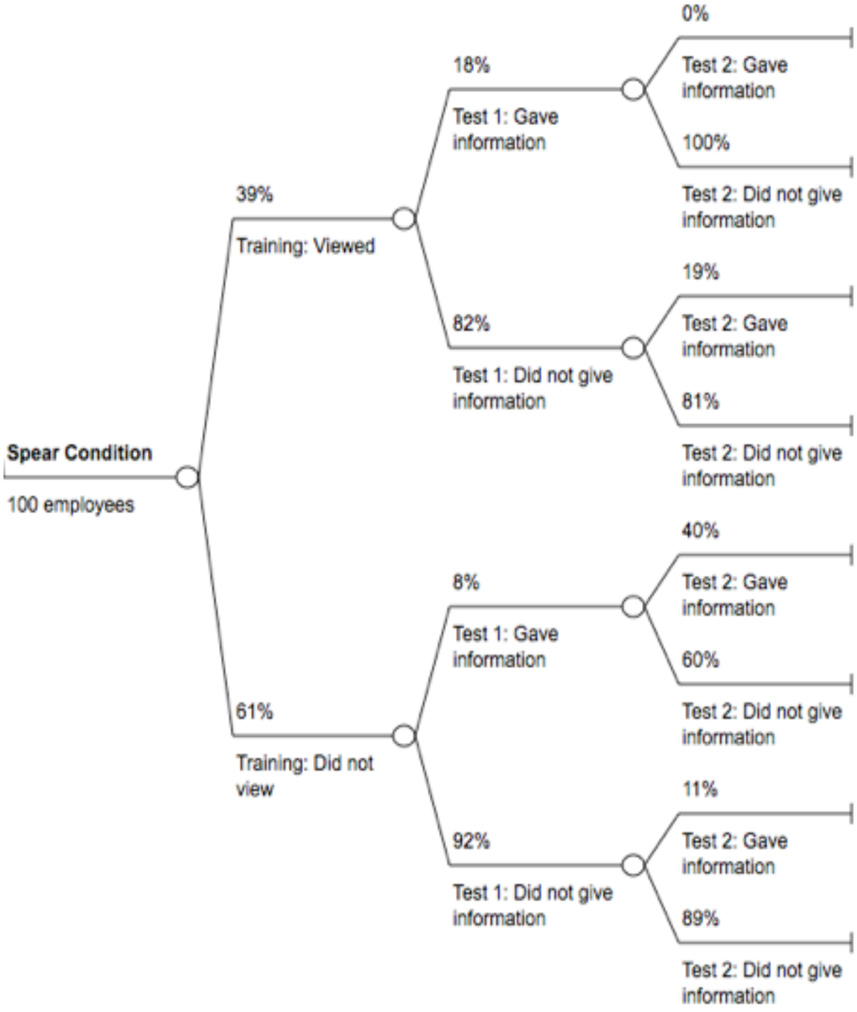


Figure 7.7: Decision tree for spear condition. It presents the percentage of employees who clicked on links in the email and gave information and the percentage of employees who did not click on links.

These results demonstrate that participants in the generic and spear training conditions were able to make better decisions immediately after being trained and that they were able to retain the knowledge for at least 7 days.

Table 7.5: Percentage of participants who clicked on link on day 0, and percentage who clicked on the link and gave information on day 2 and day 7.

	Day 0	Day 2	Day 7
Generic trained	100 %	19 %	12 %
Spear trained	100 %	18 %	15 %
Untrained	0 %	10 %	13 %

Table 7.6: Percentage of participants correctly identifying (who did not click on the link in the email) the day 2 and day 7 emails. The untrained group includes participants from both training groups who did not actually receive training.

Conditions	Identified 2 emails correctly	Identified 1 email correctly	Identified 0 email correctly
Control	58.2 %	32.8 %	8.9 %
Generic trained	70.4 %	18.5 %	11.1 %
Spear trained	65.2 %	30.4 %	4.3 %
Untrained	73.4%	22.8 %	3.8 %

Legitimate emails

We do not have enough data to conclude whether or not training increased the concern level of the participants so much that they refrained from clicking on any email links, even legitimate ones. Legitimate organizations and people often send legitimate links through emails; as such, not clicking on legitimate links may inconvenience the user. Only three employees across all three conditions clicked on the link in the legitimate email sent on day 10. To verify this behavior, we sent another legitimate email on day 14 from the marketing team, with a link to a company sales report. Again, only three employees across all conditions clicked on the link. There was no difference between the control and training (generic and spear) conditions. This suggests that the behavior may not be the effect of training, but rather the normal behavior of employees in this company towards such corporate emails.

The content of the training and testing emails used in the study has to be properly designed to provide incentives for the participants. Employees may not read email messages unless they are very relevant or involve something with severe consequences. Since we do not have enough data on how participants respond to legitimate emails, we cannot support or reject this part of Hypothesis 2. We further investigated the effect of training on legitimate emails in the second real-world study (Section 7.2). We found that training users to recognize phishing emails using PhishGuru does not make them more likely to identify legitimate emails as phishing emails.

Generic vs. spear instructions

Results suggest that the difference in the instructions in the training materials did affect participants' ability to identify phishing emails. Table 7.4 shows that the percentage of participants who clicked on the link and gave information on day 2 differed non-significantly between the generic training condition and spear training condition (two sample T-test, p-value = 0.53). The difference on day 7 was also insignificant (two sample T-test, p-value = 0.67). In Table 7.5, we examine only those participants in the generic and spear conditions who actually received training, finding that there was no significant difference between the trained conditions for the test email on day 2 (two sample T-test, p-value = 0.8) or day 7 (two sample T-test, p-value = 0.7). This suggests that participants don't gain anything by seeing specific instructions rather than generic ones.

Using both the total percentage of participants clicking on the link on day 0 (see Table 7.4) and the percentage of employees who clicked on the link on day 2 and day 7 (see Table 7.5), we found no significant difference between employees in the generic and spear conditions in their ability to identify phishing emails. Thus, we must reject Hypothesis 3. However, we believe this hypothesis warrants further investigation. A more substantial difference in the instructions between the generic and spear training might produce a significant effect. In addition, because all of the participants in this study worked on the same floor of an office building, we are concerned that participants across conditions may have shared training materials with each other. Further investigation is needed to understand the influence of interventions on decision making.

Observations

We have anecdotal evidence that employees discussed the study among themselves and with their system administrators, and we believe this had an impact on the results. Although only 50 employees clicked on the training material link, the logs show that the material was downloaded 95 times during the study (which means that some employees viewed the training material multiple times). Some people may have shown the training to colleagues in other conditions. This likely caused participants in the control condition to make correct decisions on day 2 and day 7, even though they received no direct training. However, they may have received indirect training when participants in the training conditions told them about their training or showed them the training messages. We have anecdotal evidence that employees did not receive any other information about phishing from the company and that there was no drastic incident that could have influenced employees to change their behavior during the study. This suggests that PhishGuru training can effectively train people who are not part of the study — and that, in general, it may be good enough to train a subset of employees who, in turn can influence other employees in the company. Researchers have shown that physical proximity and social structure may trigger information flow [31]. We attribute the study's result to the way the employees were seated in the company – all employees were on the same floor. Further investigation may explain this phenomenon better.

Job type did not have any influence on participants' ability to identify phishing emails before or after training. In particular, we compared technical and non-technical job types. Before the training, the percentage of participants in the generic training condition who clicked on the link and gave information was the same (42%) for technical and non-technical employees. For the spear condition, this percentage was 48% for technical and 34% for non-technical participants. This difference was statistically insignificant (two sample T-test, p-value = 0.16). Similarly, we found no significant difference between technical and non-technical employees after the training.

We found no significant difference in susceptibility to phishing emails between male and female employees (Two sample T-test, p-value = 0.76). Other researchers have found similar results [53,110,181].

We circulated a post-study questionnaire to participants to get their feedback about PhishGuru training and the training materials. Unfortunately, none of the employees turned in their completed questionnaire. We addressed this in the real-world study discussed in Section 7.2 and got a good response from the participants.

7.1.4 Discussion

The study results supported Hypothesis 1 and Hypothesis 2 (for phishing emails, with further investigation needed for the legitimate emails). Further research is needed to investigate Hypothesis 3.

In this section, we have presented the first empirical evaluation of embedded training methods that teach people about phishing during their normal use of email in the real world. We showed that in the real world: (1) a large percentage of people who click on links in simulated emails proceed to give some form of personal information; (2) PhishGuru training, a form of embedded training, effectively trains people (3) users retain knowledge for at least one week when trained with embedded training; (4) people trained with specific spear training instructions cannot identify spear phishing emails any better than people trained with generic training instructions.

Results from this study were consistent with earlier laboratory studies that demonstrated the effectiveness of the PhishGuru embedded training system. The results suggested that a large percentage of people who clicked on links in emails proceeded to give some form of personal information. We found the same results in one of our laboratory studies (Section 6.1). The results also strongly suggested that PhishGuru effectively trains employees in the real world. In earlier studies, we showed that users are more motivated to learn when training materials are presented after they fall for phishing emails (when users click on the link in the email) than when they are simply sent instructional materials through email (non-embedded). In this section, we showed that users' ability to identify phishing emails improved after training. Due to a lack of data, we were not able to conclude anything about legitimate emails in this study, but, in our next real-world study, we addressed this and were able to collect data for legitimate emails. Results also suggested that employees retain the knowledge they gain by reading training material for at least 7 days. We found similar results in earlier laboratory studies (Chapter 6). Results from this study showed that a significant number of participants identified both testing emails correctly, as compared to participants who identified one or none correctly.

Seventy nine percent of the participants clicked and gave information before training in a laboratory study, while 41% clicked and gave information in the real world. Seven days after training, this percentage fell to 35% in the lab study and 13% in the real world. The observed differences between the laboratory and real world studies may be due to a difference in participant demographics, difference in language of the study materials (English versus Portuguese), or difference in the type of simulated phishing emails used. It may also be due to the fact that real world participants use their own credentials while those in the lab use fictitious details. Despite the initial differences, participants in both the laboratory and real world study were able to learn from the training materials.

Results from this study suggested that there is no significant difference between employees trained through generic training instruction and spear training instruction when it comes to identifying phishing emails. This may be due to the small sample size of employees who were trained and who clicked on the link and gave information for the testing emails; the results also may have been influenced by employees discussing the study among themselves. Employees discussing the topic among themselves might not have been good for the study, but it does suggest that, by training a subset of employees, a company can expect these trained employees to influence other employees who were not part of the training. It would have been useful if we had more data to show this effect, but this may be a good starting point for further investigation on the topic.

7.2 Second evaluation: Carnegie Mellon University

This section is largely a reproduction of a paper co-authored with Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Ted Pham. It is under review at a conference [108].

Based on the limitations and lessons learned from the study discussed in Section 7.1, we designed a field trial with Carnegie Mellon University; for this study, we had better control over the experiment and collected richer data. Details and results of the study are discussed in this section.

The remainder of the section is organized as follows: in Section 7.2.1, we discuss the recruitment of participants and participant demographics. In Section 7.2.2, we present the study setup and in Section 7.2.3, we present the hypotheses that guided the study. In Section 7.2.4, we present the results of the evaluation, demonstrating that PhishGuru effectively educates people in the real world. In Section 7.2.5, we present the challenges of conducting a field trial to study the effectiveness of phishing interventions and the ways in which we addressed them. Finally, in Section 7.2.6, we discuss the effect of training people in the real world.

7.2.1 Recruitment and demographics

We sent a recruitment email to all active CMU students, faculty, and staff Andrew email accounts[1] with the primary campus affiliation listed as "Pittsburgh." The email subject line read "Volunteers Needed: Help Us Protect the Carnegie Mellon Community from Identity Theft," and the email content described both what would be required of participants and what data would be collected from them. In addition, they were told that volunteers would be entered into a raffle to receive one of five $75 gift cards. Willing participants were instructed to reply to the recruitment email or go to a web link to opt in to the study. We also added "To verify the authenticity of this message, visit the ISO[2] Security News & Events at https://www.cmu.edu/iso" so that users could check the legitimacy of the message. In total, we sent 21,351 emails and recruited 515 volunteers. The Human Resources department at CMU provided us with the participant demographics presented in Table 7.7.

Every person in the university was assigned a primary department, even if they were students with double-majors or faculty with joint appointments. For the purpose of this study and analysis, we looked only at their primary departments (listed as department in Table 7.7). As shown in Table 7.7 we grouped the 26 different departments into 7 academic department clusters and 3 non-academic department clusters. For example, we grouped the Entertainment Technology Center and School of Computer Science together as Computer Science.

[1] The Andrew account is the main email account given to all CMU community members.

[2] Information Security Office at CMU.

Table 7.7: Percentage of people in the three conditions and percentage of people who fell on day 0 in each demographic (N = 515).

	% of control	% of one-train	% of two-train	% who fell for day 0 phish
Gender				
Female	44.8	48.8	39.8	48.5
Male	55.2	51.2	60.2	50.7
Affiliation				
Faculty	7.0	8.7	7.0	38.5
Staff	36.0	38.4	30.4	37.8
Students	56.4	52.9	62.6	58.6
Sponsored	0.6	0	0	0
Student year				
Doctoral	13.4	17.5	12.3	52.7
Masters	19.8	19.8	21.7	56.2
Undergraduate	20.9	18.6	28.0	62.9
Miscellaneous	2.3	1.1	0	66.7
None	43.6	43.0	38.0	37.9
Department type				
Academic	72.7	73.9	78.4	53.1
Administrative	24.4	24.4	19.3	39.3
Unknown	2.9	1.7	2.3	41.7
Academic departments				
IS and Public Policy	8.7	12.2	12.8	50
Humanities & Social Sciences	7.6	8.7	8.1	59.5
Engineering	16.3	14.5	14.6	57.7
Fine Arts	4.6	6.4	3.5	48
Computer Science	16.3	14.5	18.7	48.2
Business	8.7	5.8	10.5	51.2
Sciences	10.5	11.6	11.1	52.6
Non-academic departments				
Computing Services and Research	5.8	5.8	5.2	34.5
Administration	18.6	18.0	13.6	41.2
Other	2.9	2.3	1.8	50

Table 7.8: Schedule of the emails including day of study, calendar date (2008), and type of emails sent out that day. For example, on day 0, we sent test and legitimate emails to all participants.

Study day	Day 0	Day 2	Day 7	Day 14	Day 16	Day 21	Day 28	Day 35
Date	Nov 10	Nov 12	Nov 17	Nov 24	Nov 26	Dec 1	Dec 8	Dec 15
Type of Emails Sent	Train and test, then legitimate	Test	Test, then legitimate	Train and test	Test	Test	Test, then legitimate	Post-study survey

7.2.2 Study setup

Five hundred and fifteen participants were randomly assigned to three conditions: "*control*," "*one-train*," and "*two-train*." There were 172 participants in control, 172 in one-train, and 171 in two-train. As shown in Table 7.8, all participants, regardless of condition, were sent a series of 3 legitimate and 7 simulated spear-phishing emails over the course of 28 days. In the body of each email was a simulated phishing URL. Clicking on this link resulted in different scenarios depending on the study day and the participant's condition. Participants in the one-train condition who clicked on the URL on day 0 and participants in the two-train condition who clicked on the URL on day 0 and/or day 14, saw one or both (one on each day) of the anti-phishing training interventions depicted in Figure 7.8. For all other study days in the one-train and two-train conditions, clicking on the URL led to a simulated phishing webpage where an HTML form asked users to provide private credentials. Participants in the control condition did not receive any anti-phishing training as part of the study. When they clicked on the URLs, they were directed to simulated phishing webpages. We tested participants twice after each training email for immediate retention (2 days) and short-term retention (7 days). This data also helped us confirm the immediate and short-term retention results from earlier studies (laboratory and real-world).

Table 7.9 presents an overview of the 7 simulated phishing emails sent to participants. Except for the "Community Service" email—which proved to be a much less effective phishing lure than the other messages—we found no difference in the rate at which participants fell for each of the emails on day 0. However, to ensure that the aggregate response rates per day were not confounded by the potential difference in natural response rates for individual emails, or by the interdependence of response rates among the emails, we developed a counterbalancing schedule. The counterbalancing schedule avoided these confounding issues by dividing the 515 participants randomly and equally per condition among 21 different viewing schedules for the 7 emails. The critical property of the 21 schedules was that, for any given day of the study, each of the 7 emails was sent out to an equal number of participants. This allowed us to compute the aggregate response rate for an entire day by summing the responses to each of the emails sent that day. Since the proportions were constant

Table 7.9: Summary of emails sent to study participants. In all emails, when the user clicked on the link in the email, she was taken to a page where her user name and password was requested. The "Bandwidth Quota Offer" email gave users an opportunity to increase their daily wireless bandwidth limit. The "Plaid Ca$h" email contained instructions to claim $100 in Plaid Ca$h (money to be used at CMU vendors). The remaining emails are sufficiently explained by the subject line. The legitimate email had "https" while all others had "http" in the URL.

Email type	**Subject line**	**Domain name in URL**
Test/Train	Bandwidth Quota Offer	cmubandwithamnesty.org
Test/Train	Register for Carnegie Mellon's annual networking event	carnegiemellonnetworking.org
Test/Train	Change Andrew password	andrewpasswordexpiry.org
Test/Train	Congratulation - Plaid Ca$h	idcardsforcmu.org
Test/Train	Please register for the conference	studenteventsatcmu.org
Test/Train	Volunteer at Community Service Links	communityservicelinks.org
Test/Train	Your Andrew password alert	andrewwebmail.org
Legitimate	Earn Bonus Points #1: Win a Nintendo Wii, $250 Amazon Gift Card or other great prizes	cmu.edu

for all study days, different aggregate response rates across different days were comparable. To counterbalance the training materials, half of the participants in the one-train condition received intervention A (see Figure 7.8) and the other half received intervention B (see Figure 7.8). Similarly, in the two-train condition, half of the participants received intervention A first and intervention B second while the other half received intervention B first and intervention A second. We found no significant difference in response rates among participants who received the training materials in different orders or among those who received different training material.

All emails constructed for the study were emails that the CMU community might normally receive, though they were not based on any information that a phisher would be unable to obtain from public webpages. Based on the headers of the email messages participants sent us to sign up for the study, we determined that a large fraction of the participants used Squirrel Mail, which by default strips HTML from email messages. Therefore, we did not replicate the common phishing tactic of using HTML to hide phishing URLs from users. All of the phishing messages displayed the phishing URLs in the body of the messages. Figure 7.9 (Top) shows an example of an email that was used in the study. This particular example asks the study participant to click on the link to change their Andrew password.

We registered all of the domain names in the simulated phishing emails using legitimate credentials (Table 7.9)—that is, a query to the associated "whois" database would show valid CMU affiliated contact information. In this way, if participants were skilled enough, they could easily infer that these domains were part of the study. Besides those shown in Table 7.9, we also registered another

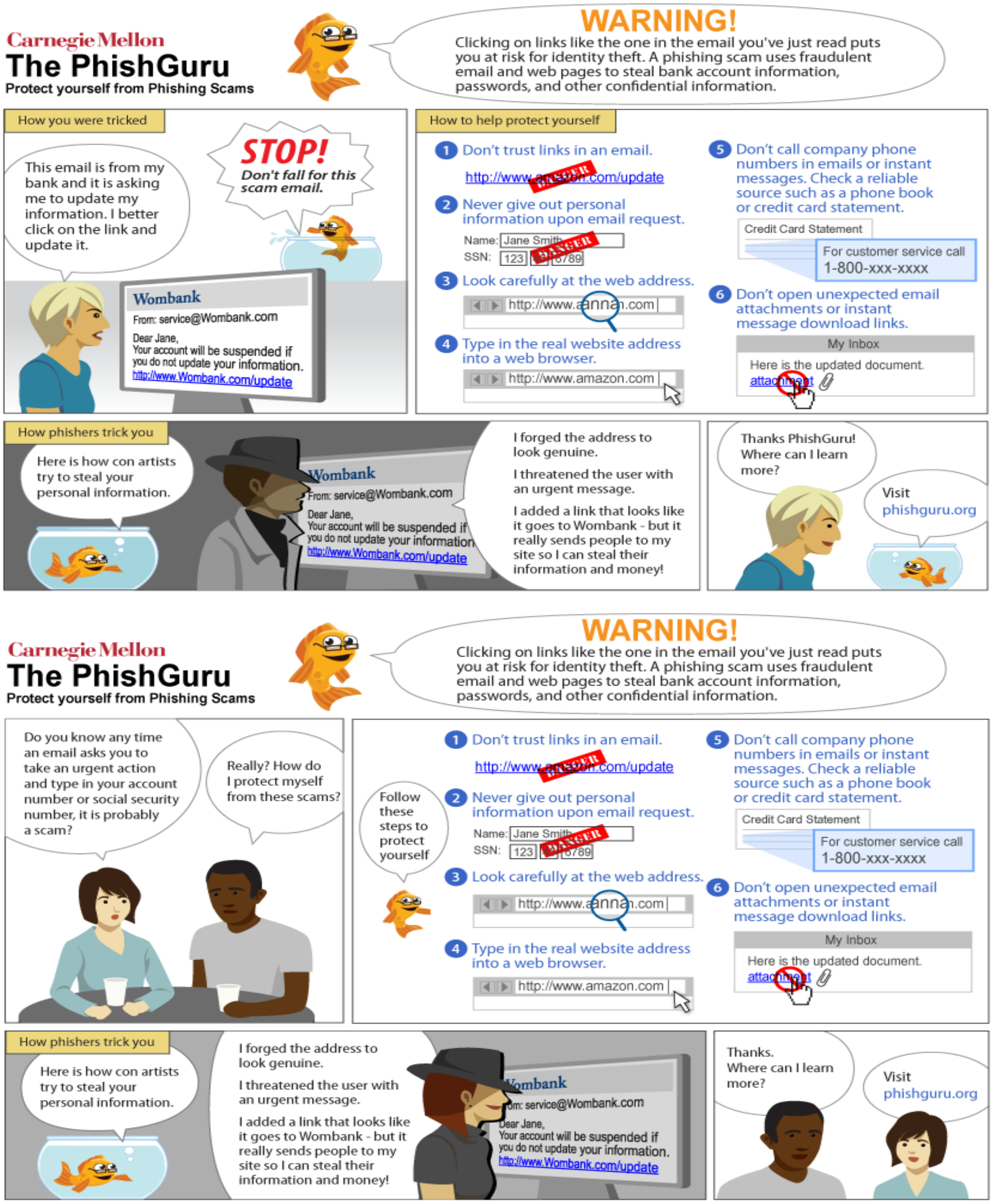
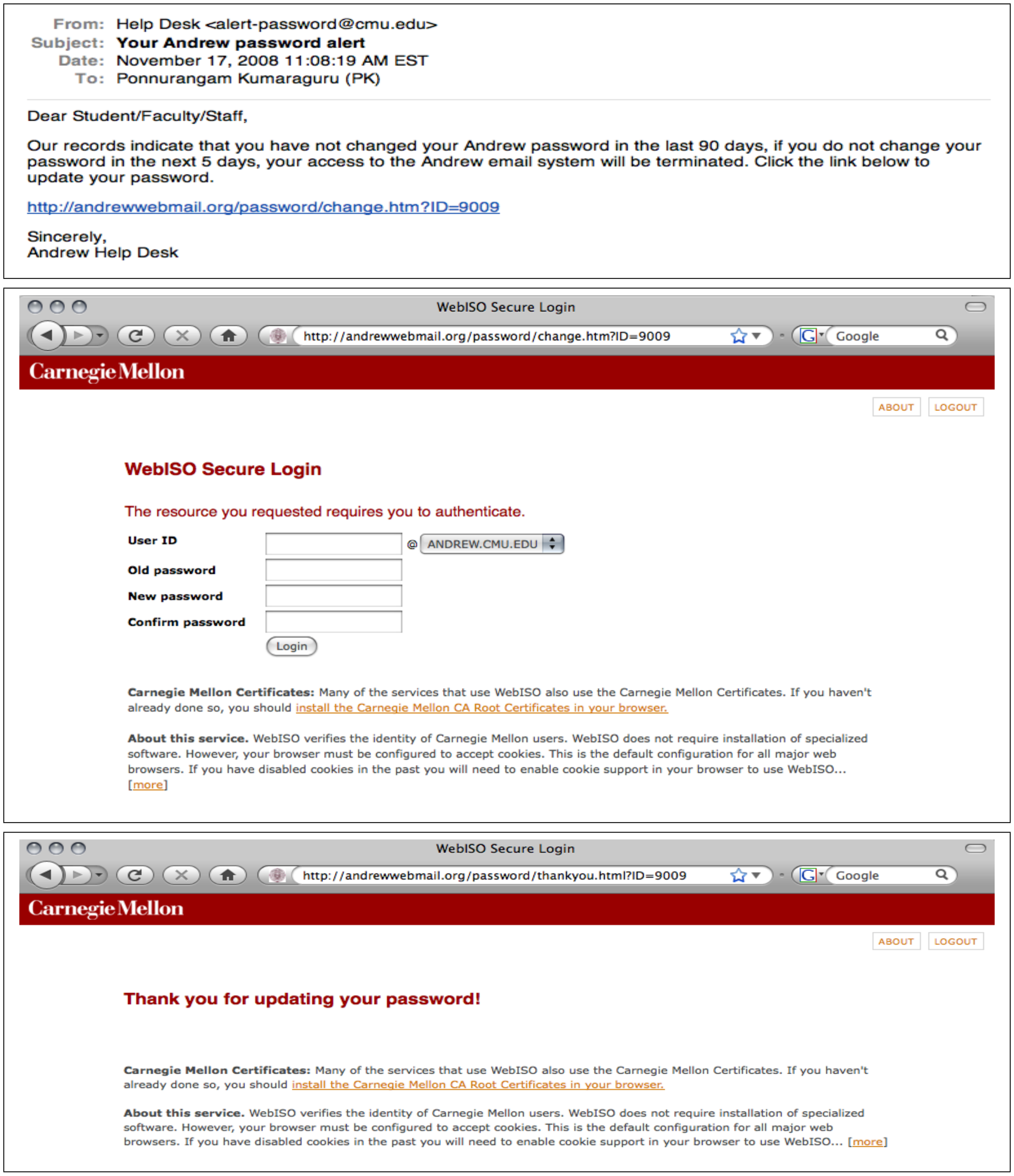

Carnegie Mellon
The PhishGuru
Protect yourself from Phishing Scams
WARNING!
Clicking on links like the one in the email you've just read puts you at risk for identity theft. A phishing scam uses fraudulent email and web pages to steal bank account information, passwords, and other confidential information.
How you were tricked
This email is from my bank and it is asking me to update my information. I better click on the link and update it.
STOP! Don't fall for this scam email.
Wombank
From: service@Wombank.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.Wombank.com/update
How to help protect yourself
1 Don't trust links in an email.
http://www.amazon.com/update
DANGER
2 Never give out personal information upon email request.
Name: Jane Smith
SSN: 123 45 6789
DANGER
3 Look carefully at the web address.
http://www.aannan.com
4 Type in the real website address into a web browser.
http://www.amazon.com
5 Don't call company phone numbers in emails or instant messages. Check a reliable source such as a phone book or credit card statement.
Credit Card Statement
For customer service call 1-800-xxx-xxxx
6 Don't open unexpected email attachments or instant message download links.
My Inbox
Here is the updated document.
attachment
How phishers trick you
Here is how con artists try to steal your personal information.
Wombank
From: service@Wombank.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.Wombank.com/update
I forged the address to look genuine.
I threatened the user with an urgent message.
I added a link that looks like it goes to Wombank - but it really sends people to my site so I can steal their information and money!
Thanks PhishGuru! Where can I learn more?
Visit phishguru.org
Carnegie Mellon
The PhishGuru
Protect yourself from Phishing Scams
WARNING!
Clicking on links like the one in the email you've just read puts you at risk for identity theft. A phishing scam uses fraudulent email and web pages to steal bank account information, passwords, and other confidential information.
Do you know any time an email asks you to take an urgent action and type in your account number or social security number, it is probably a scam?
Really? How do I protect myself from these scams?
Follow these steps to protect yourself
1 Don't trust links in an email.
http://www.amazon.com/update
DANGER
2 Never give out personal information upon email request.
Name: Jane Smith
SSN: 123 45 6789
DANGER
3 Look carefully at the web address.
http://www.aannan.com
4 Type in the real website address into a web browser.
http://www.amazon.com
5 Don't call company phone numbers in emails or instant messages. Check a reliable source such as a phone book or credit card statement.
Credit Card Statement
For customer service call 1-800-xxx-xxxx
6 Don't open unexpected email attachments or instant message download links.
My Inbox
Here is the updated document.
attachment
How phishers trick you
Here is how con artists try to steal your personal information.
I forged the address to look genuine.
I threatened the user with an urgent message.
I added a link that looks like it goes to Wombank - but it really sends people to my site so I can steal their information and money!
Wombank
From: service@Wombank.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.Wombank.com/update
Thanks. Where can I learn more?
Visit phishguru.org

Figure 7.8: Above: Intervention A. One of the two training interventions used in the study. One half of the participants in the one-train and two-train conditions received this training intervention on day 0. The other half of the two-train condition received this on day 14. Below: Intervention B. The second training intervention used in the study. The instructions are the same as in Intervention A, but the characters and the story are slightly different. One half of the participants in the one-train and two-train conditions received this training intervention on day 0. The other half of the two-train condition received this on day 14.



From: Help Desk <alert-password@cmu.edu>
Subject: **Your Andrew password alert**
Date: November 17, 2008 11:08:19 AM EST
To: Ponnurangam Kumaraguru (PK)

Dear Student/Faculty/Staff,

Our records indicate that you have not changed your Andrew password in the last 90 days, if you do not change your password in the next 5 days, your access to the Andrew email system will be terminated. Click the link below to update your password.

http://andrewwebmail.org/password/change.htm?ID=9009

Sincerely,
Andrew Help Desk

WebISO Secure Login

The resource you requested requires you to authenticate.

User ID @ ANDREW.CMU.EDU

Old password

New password

Confirm password

Carnegie Mellon Certificates: Many of the services that use WebISO also use the Carnegie Mellon Certificates. If you haven't already done so, you should install the Carnegie Mellon CA Root Certificates in your browser.

About this service. WebISO verifies the identity of Carnegie Mellon users. WebISO does not require installation of specialized software. However, your browser must be configured to accept cookies. This is the default configuration for all major web browsers. If you have disabled cookies in the past you will need to enable cookie support in your browser to use WebISO... [more]

Thank you for updating your password!

Carnegie Mellon Certificates: Many of the services that use WebISO also use the Carnegie Mellon Certificates. If you haven't already done so, you should install the Carnegie Mellon CA Root Certificates in your browser.

About this service. WebISO verifies the identity of Carnegie Mellon users. WebISO does not require installation of specialized software. However, your browser must be configured to accept cookies. This is the default configuration for all major web browsers. If you have disabled cookies in the past you will need to enable cookie support in your browser to use WebISO... [more]

Figure 7.9: A sample of simulated phishing emails and websites. Top: A sample of the simulated phishing emails used in the study. The URL that appears in the email matches the target of the HREF statement. Middle: One of the seven simulated websites. Using JavaScript, all of the form data the user submitted was discarded prior to form submission. Bottom: "Thank you" webpage shown to the users when they gave credentials on the webpage presented in Middle. Similar pages were presented for other simulated websites.

10 similar-looking domains as backup.

Figure 7.9 (Middle) shows one of the simulated phishing websites. This example simulates the standard password change scenario at CMU. The site asks participants to provide their User ID, old password, and new password, and then to confirm their new password. All of the websites used in the study collected some combination of user name and password in a similar fashion. As shown in Figure 7.9 (Bottom), when participants submitted their information, they were taken to a "thank you" page. Participants saw a similar sequence of webpages ("login" followed by "thank you") in all email scenarios.

To estimate the false positive rate, we measured the response rate to three legitimate emails sent to study participants by the CMU Information Security Office (ISO). These messages were sent to all participants on day 0, day 7, and day 28 after the test/training emails were sent. The original recruitment email for this study was presented in the context of Cyber Security Awareness Month. The three legitimate emails were announcements for an ongoing security related scavenger hunt which had begun during Cyber Security Awareness Month and gave community members an opportunity to gain points in return for specified security related tasks. The subject line of the first email was "Earn Bonus Points #1: Win a Nintendo Wii, $250 Amazon Gift Card or other great prizes." The second and third emails had identical subjects, except that they were emails "#2" and "#3," respectively. The email itself indicated that the recipient needed to login with their Andrew password to claim their bonus points. Clicking the link took them to the real "webiso login page" (the standard log-in page for all CMU websites—the one that we spoofed in the phishing websites), where they were asked to provide their username and password.

In order to track user responses, each participant was given a unique 4-character alpha-numeric hash that was appended as a parameter to the URL of all emails participants received (*e.g.* in one email, participant 9009 received a URL that ended with `update.htm?ID=9009`). The hash also served as a mechanism to allow us to protect the identity of participants during data analysis. To ensure that no sensitive data would be compromised, ISO did a complete penetration test on the machine that was used to host the phishing websites. In addition, the simulated phishing webpages were constructed so that no information was ever submitted to the webserver. Using JavaScript, all of the form data the user submitted was discarded prior to form submission. To ensure that the emails were not blocked by CMU spam filters, the machine from which the emails were sent was put on a white list.

After all real and simulated phishing emails were sent, another email was sent to all participants asking them to complete a post-study survey. The survey consisted of questions regarding: (1) the interest level of participants in receiving such training in the future; (2) participants' feedback on the training methodology; (3) participants' feedback on the interventions and instructions; (4) whether participants remembered registering for the study; and (5) demographic information such

Table 7.10: Percentage of participants who clicked and gave information on days 0 through 28. N is the number of participants in each condition. Participants in the training conditions saw the interventions on day 0 and therefore it is NA (not applicable) in the "G" column. We found no significant differences among the click rates of participants across the three conditions on day 0 and among participants in the control group on all days. C means "clicked" and G means "gave."

Conditions	N	Day 0		Day 2		Day 7		Day 14		Day 16		Day 21		Day 28	
		C	G	C	G	C	G	C	G	C	G	C	G	C	G
Control	172	52.3	40.1	51.2	39.5	48.3	40.7	54.1	41.3	44.1	30.8	41.3	25.0	44.2	30.8
One-train	172	51.7	NA	35.5	29.1	34.9	26.7	35.5	25.0	23.8	19.2	29.7	22.1	23.8	17.4
Two-train	171	45.0	NA	31.6	23.9	30.4	21.6	37.4	NA	29.2	21.6	26.9	18.1	25.6	17.5

as age. Two hundred and seventy nine of the participants completed the post-study survey. These participants were distributed nearly equally across the three conditions (control = 31.5%; one-train = 34.0%; two-train = 34.5%).

7.2.3 Hypotheses

In this section, we describe the hypotheses tested in this study. The goal in this study was to investigate whether PhishGuru helps people retain long term knowledge about phishing. In particular, the aim was to study retention after 28 days.

> **Hypothesis 1**: Participants in the training conditions (one-train and two-train) identify phishing emails better than those in the control condition on every day except day 0.

Earlier studies only tested the effectiveness of the training methodology when participants were trained once. Learning science literature however, suggests that if people are provided with more opportunities to learn, they tend to better remember instructions [59]. In PhishGuru, the simulated email works for both training and testing purposes; people who continue to click on the simulated phishing URLs can be presented with further training materials. The goal was to investigate whether participants who read the training materials twice had any advantage over participants who read the training materials only once.

> **Hypothesis 2**: Participants who see the training interventions twice perform better than participants who see the intervention once.

Earlier studies did not provide any conclusive evidence for whether training has any effect on false positive errors (Section 7.1). We believe that it is very important to consider this criterion when measuring training success. In this study, we sent legitimate emails to participants on day 0, day 7, and day 28 to measure the false positive error rate.

Hypothesis 3: When asked to identify legitimate emails, participants who view the training materials in the training conditions will perform the same as participants in the control condition.

7.2.4 Results

In this section, we present the results of the study. The results of this study support Hypotheses 1, 2, and 3. In this study, we did not use any type of web bug or other method to determine how many participants in the CMU study failed to click on links because they never opened the email. However, we would expect this behavior to occur at similar rates across all conditions, so it does not impact our conclusions.

Long-term retention

Results show that people in the one-train and two-train training conditions who fell for the first phishing message performed significantly better when they received the second phishing message than those in the control condition. In addition, we observed no significant loss in retention after 28 days. Table 7.10 presents the percentage of participants who clicked and gave information on day 0 through day 28. Approximately 52.3% (90 participants) in control, 51.7% (89 participants) in one-train, and 45.0% (77 participants) in the two-train conditions clicked on the link in the email they received on day 0. We found no significant difference among the click rates of participants across the three conditions on day 0 (ANOVA, F(2,512) = 1.1, p-value = 0.3). This implies that, prior to any influence from the study, participants in all three conditions were similar. We also found no significant difference (ANOVA, F(6,1203)= 1.7, p-value = 0.3) in the click rate of participants in the control group across study days (day 0 until day 28). This implies that there was no change in the behavior of participants in the control group throughout the study.

On day 0, 48.4% of the participants in the training conditions viewed the PhishGuru intervention. To determine the effectiveness of the training, we conditioned the click rates of days 2 through 28 on those participants across all conditions who clicked on the links in the email(s) on day 0. This allowed us to compare the participants who actually received the training in the one-train and two-train conditions to those in the control condition who took the analogous action on day 0. Figure 7.10 (Left) shows the percentage of these participants who clicked on links in emails and gave information to the fake phishing websites from day 2 until day 28. There is a significant difference (Chi-Sq = 14, p-value < 0.001) between the percentage of users who clicked in the control condition (54.4%) and the percentage who clicked in the one-train condition (27.0%) on day 28. Similarly, there is significant difference between the control and two-train (32.5%) conditions on day 28 (Chi-Sq = 8.9, p-value < 0.01). We also found that, in the one-train condition, participants who gave information to fake phishing websites on day 2 were not significantly different on day 28 (Chi-Sq = 3.5, p-value < 0.1). Similarly, there is significant difference between the control and one-train conditions and between the control and two-train conditions in the percentage of people who clicked on days 2 through 28. This shows that users trained with PhishGuru retain knowledge even after 28 days, supporting Hypothesis 1.

Multiple training

Results strongly suggest that users who saw the training intervention twice were less likely to give information to the fake phishing websites than those who only saw the training intervention once. Figure 7.10 (Right) shows the percentage of participants who clicked on links in emails from day 16 until day 28 conditioned on participants who clicked on the link on day 0 and those who clicked on day 14. There is a significant difference (Chi-Sq = 5.4, p-value =0.01) between the percentages of users who clicked in the one-train condition (42.9%) and those who clicked in the two-train condition (26.5%) on day 16, and a similar difference on day 21 (Chi-Sq = 7.8, p-value < 0.01). However, we did not find a significant difference between users who clicked in the one-train and two-train conditions on day 28 (Chi-Sq = 0.3, p-value =0.6). In the tow-train condition, we also did not find any significant difference (Chi-Sq = 1.1, p-value = 0.3) in clicking between day 21 (26.5%) and day 28 (35.3%).

Figure 7.10 (Right) also shows that participants who were trained twice did significantly better than those who were trained once when it came to giving their personal information to fake phishing websites. For example, on day 28, 31.4% of the participants in the one-train condition gave information to the website, while only 14.7% did in the two-train condition. This is significantly different (Chi-Sq = 7.3, p-value < 0.01), supporting Hypothesis 2.

We also found that 30 participants (17.5%) in the two-train condition who did not see the intervention on day 0 saw the intervention on day 14. These are the people who probably needed training, since they fell for the email on day 14. We saw no significant difference (t-test, t = 0.1, p-value = 0.8) between people in the one-train condition who clicked on day 14 but were trained on day 0 and people in the two-train condition who clicked on day 28 but were trained only on day 14. This suggests that multiple rounds of training is useful not only for re-inforcement, but also for providing an additional opportunity for people who need training.

Legitimate emails

Results from this study indicate that training users to recognize phishing emails using PhishGuru does not make them more likely to identify legitimate emails as phishing emails. Table 7.11 presents the percentage of participants who clicked and gave information in response to legitimate emails

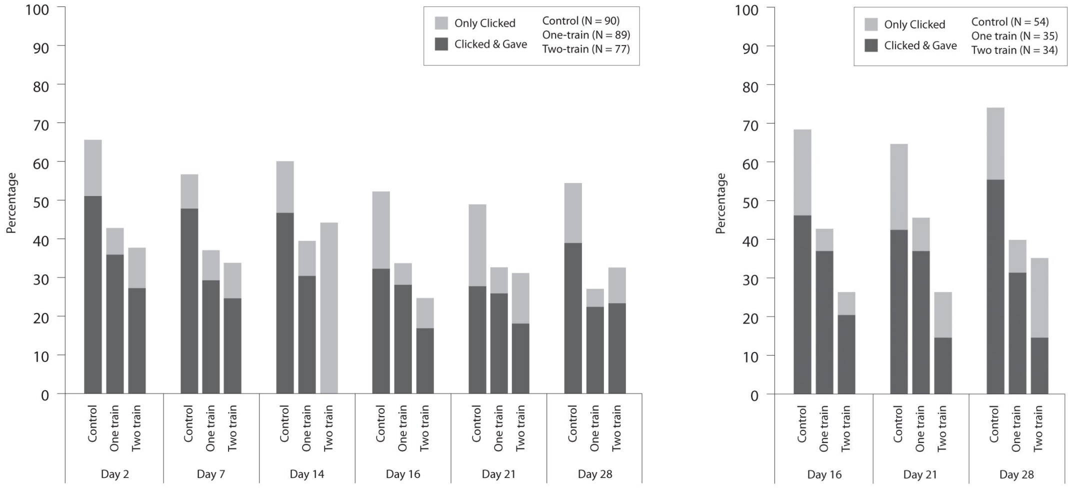


Figure 7.10: Percentage of participants who clicked on phishing links and gave information. Left: Days 2 through 28 conditioned on those participants who clicked the link on day 0. N is the number of people who clicked on day 0. Nobody gave information in the two-train on day 14 because it was a training email. There is significant difference between the control and one-train and between the control and two-train conditions in the percentage of people who clicked on days 2 through 28. Right: Days 16 through 28 conditioned on those participants who clicked on both day 0 and day 14. N is the number of people who clicked on day 0 and on day 14. There is significant difference between the one-train and two-train conditions in the percentage of people who gave information to phishing sites on days 16 through 28.

out of those participants who clicked on day 0. We found no significant difference among the three conditions on day 0 (ANOVA, F(2,512) = 2.7, p-value = 0.1). Similarly there was no significant difference among the three conditions on day 7 and day 28. Since the legitimate email used in the study was same on all three days, as expected, we see a natural decline in response rate over the course of the study (Table 7.11). This shows that user behavior did not change with respect to the legitimate emails tracked as part of the study, confirming that training people does not decrease their willingness to click on links in legitimate email messages. This result supports Hypothesis 3.

Analysis based on demographics

Multivariate regression analysis did not find any significant relationship between susceptibility to phishing on day 0 and gender (p-value = 0.9 for gender coefficient), student year (p-value = 0.5 for student year coefficient), or department (p-value = 0.8 for department coefficient). However, we did find significant difference based on affiliation. In particular, we found significant difference (Std. error = 0.2, p-value < 0.05) between students and staff in falling for phishing on day 0. We found that students were more vulnerable to phishing emails before receiving any training from the study. We also found significant difference in the department type (different from primary department). In particular we found significant difference (Std. error = 0.2, p-value < 0.05) between the academic

Table 7.11: Percentage of participants who clicked and gave information to the legitimate emails out of those participants who clicked on day 0. N is the number of participants in each condition. There is no significant difference between the three conditions on any given day.

Condition	N	Day 0		Day 7		Day 28	
		Clicked	Gave	Clicked	Gave	Clicked	Gave
Control	90	50.0	42.2	41.1	37.8	38.9	35.6
One-train	89	39.3	38.2	42.7	37.1	32.3	30.3
Two-train	77	48.1	36.3	44.2	36.4	35.1	32.5

and administrative department types, with academics being more susceptible to falling for the phishing email. Investigating further, we found that the difference could be attributed to the fact that all students are in the academic department type, making this group as a whole more vulnerable than others.

We investigated this difference between students and staff further to see if age was a factor in susceptibility to phishing. We used the age data collected through post-study surveys. Two hundred and sixty-seven participants provided their age in the survey. The minimum age in years was 18 and the maximum age was 77 (avg. = 32.3, SD = 12.8). We found a significant difference (Chi-Sq = 8, p-value < 0.01) in the likelihood of clicking on links on day 0 between 18 - 25 age group and those in all of the older age groups (Shown in Table 7.12). This shows that, prior to any training, those participants in the 18-25 age group are more likely to click on links in phishing emails than any other age group.

Among the participants who were trained on day 0, again, multivariate regression analysis did not find any significant relationship between susceptibility to phishing on day 28 and gender (p-value = 0.4 for gender coefficient), student year (p-value = 0.9 for student year coefficient), and department (p-value = 0.7 for department coefficient). We did find difference (Std. error = 0.3, = p-value $<$ 0.001) between the academic and administrative department types, which was again attributable to students falling for phishing after training. As with day 0, on day 28 we found that the age group 18 - 25 was significantly (Chi-Sq = 10.5, p-value < 0.01) more likely to fall for phishing than other age groups (Table 7.12). We found that participants in the 18-25 age group were consistently more vulnerable to phishing attacks on all days of the study than older participants. These results are in line with risk averse literature, which says that younger people are more likely to be impulsive, while older people are risk averse and less impulsive [147]. We were not able to draw any concrete conclusions about faculty because the sample sizes were too small.

Computer savvy technical people (Software Engineering Institute, Computing Services) were less likely than others to fall for phishing emails. In general, however, participants in Computer Science and Computing Services and Research department clusters did not perform significantly differently than participants in any other group on day 0.

Table 7.12: Percentage of participants who clicked on the link in the emails by age group. N = 267 people responded to the post-study survey with their age. These results show that the 18 - 25 age group behaves in a significantly different way from all of the other age groups.

Age group	Day 0	Day 28
18 - 25	62.3	35.7
26 - 35	47.5	15.8
36 - 45	33.3	18.2
46 and more	42.5	10

Observations

In this section, we describe the data collected in the study and through the post-study survey, as well as other observations from the data.

Results indicate that any participant who will eventually click on the link in an email will do so within 8 hours from the time the email is sent. To estimate the distribution of how long people took to read emails, we used the time at which a participant clicked on the phishing link as a proxy for the time the email was read. Figure 7.11 presents the cumulative number of emails that were clicked on for each study day from the day the study email was sent out. This shows that, 2 hours after the emails were sent, at least half of the people who eventually clicked on the link had already done so; after 8 hours, nearly all people (90%) who clicked had already done so. This suggests that anti-phishing methods that rely on black-lists should aim to update their lists before this window has passed; otherwise, users will click on the link and become a victim for phishing. This further supports the effectiveness of methodologies such as PhishGuru that work from the start of a phishing attack.

Some of the post-study survey questions were designed to gauge the receptiveness of the CMU community to PhishGuru training. Participants generally liked the idea of conducting such campus studies at regular intervals (Question 1 in Table 7.13). One participant wrote, “I really like this study, and I should have this kind of program every year to increase the awareness.” Another wrote, “This should be one of the first things that incoming CMU students learn.” Some participants liked the idea of being reminded of the instructions periodically. One participant wrote, “It is always good to be reminded. Sometimes you forget, so I think getting reminders once a month is a good way of helping us to remember.” Table 7.13 (Question 2) also shows that few participants were unwilling to recommend such training to their friends. We were also interested in finding out how often the emails should be sent to the participants. We asked, “How often would you like to receive educational materials like this picture(s) in your email inbox?” Eighty five participants responded to the question; forty percent answered “Once a month,” while 22.3% said they would never want to see such training emails.

When asked to give an open-ended comment about the study, one of the participants said “One

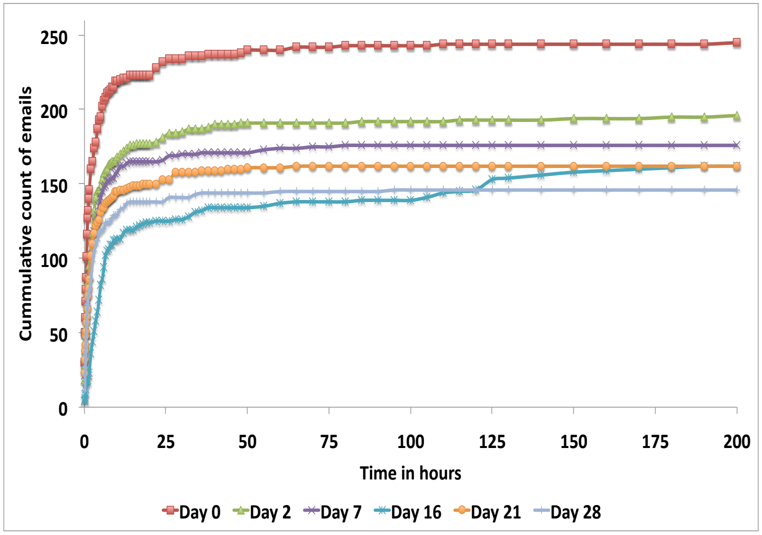


Figure 7.11: Cumulative number of emails clicked since the email was sent out. This shows that study participants who clicked on the links in emails did so within 8 hours of the time the email was sent out. Because of a technical error, we were not able to capture the data for day 14. The day 16 time-window spans the Thanksgiving holiday, with the second peak coinciding with the Monday after Thanksgiving.

Table 7.13: Post study questions. Participants enjoyed receiving training materials and recommended that CMU perform such studies regularly. N = 85.

Questions/responses	**Response in %**
(1) Would you recommend that CMU continue doing this sort of training or study in the future?	
Yes	80
Not sure	17.6
No	2.4
(2) How likely are you to recommend this type of training to a friend?	
Definitely	38.8
Maybe	51.8
Will not	9.4

thing I did not like about the study is that I was tricked by one email that was part of the study, but I had to call to be reassured that I did not have to change my Andrew password." Since we were working with the ISO team, they presented a canned response to inquiries from participants. We believe this mitigated potential backlash to the study. We also believe that, when it comes to training emails, participants who click on the link should be quickly and courteously alerted to the fact that they have been tricked. We incorporated such friendly alerts into the training messages. In the case of testing emails, it is important to debrief people about the study and provide them with opportunities to give their feedback. In the study, we debriefed participants through an email and plan to conduct a university-wide presentation about the results.

Unlike the previous PhishGuru field study, we found little interaction between participants discussing the study. Only 13% of participants indicated that they had talked about the tips presented in the PhishGuru training with other members of the CMU community in the last 30 days. Six of the participants who said they had discussed the training provided information about their discussions. A typical response was: "Just talked about the fact that I fell for one scam that offered $100 prize" or "I did talk about how I was tricked VERY easily into giving away my username/password to my andrew account." To further understand potential contamination across study conditions, we asked "How did you get to see the picture(s)?" in the post-study. Of those who responded, 87% reported seeing the training cartoons through a link in an email from the study. Only 5% reported seeing the training through a link in an email that was forwarded by a friend or a colleague at CMU, and 5% reported that a friend or a colleague at CMU showed them the training. The remaining participants said they couldn't remember how they got to the training. These results show that most of the participants received the training material through the emails sent through the study; therefore, there was little chance for interaction among participants regarding the study, and so little chance of the conditions being contaminated.

Some participants were interested in knowing more about phishing. Our log files indicate that 12.2% of the participants who got trained on day 0 (one-train and two-train) visited phishguru.org, the website that Phishguru cites as a source for more information in the intervention (Figure 7.8).

Analyzing PhishGuru using a formal framework

We also used Cranor's Human-in-Loop framework to analyze the PhishGuru training intervention. Cranor developed a framework that accounts for the role of humans in any security system [50]. This framework can be used to analyze different communication devices (warning, notice, status indicator, policy, and training). Researchers have performed similar analyses to study the effectiveness or ineffectiveness of their secure email system [124]. The framework includes many components to analyze the security system and observe human behavior, the system itself, and the way humans and systems work together. Researchers use this framework to determine which aspects of a particular system is good and which are bad.

Using this framework to analyze PhishGuru, we identified the areas where PhishGuru faces significant challenges. We inferred that organizations cannot send fake phishing emails to their customers. Given the fair business practices that organizations have to follow and their relationship with customers, they do not have the liberty and flexibility to train their customers by sending them fake phishing emails. These organizations may be ready to do this type of training with their employees, but not with their customers. However, the lack of willingness to send emails to customers is more prevalent in the US than other countries. We have heard anecdotally that organizations in other countries, where customer privacy awareness is not as prevalent, have been willing to send such training materials to their customers.

From the studies discussed in this thesis, we also found that people are able to process the information presented by PhishGuru and apply that information to future scenarios that are similar or quite different. Results from the studies also showed that people trained by PhishGuru behave differently and are able to identify phishing emails much better after PhishGuru training than they could before. From the analysis, it appears that the training is most effective when presented in a setting where the most possible people will see it [70]. In this case, PhishGuru is the ideal set-up, making perfect use of the"teachable moment" – users see the PhishGuru intervention right after they click on a link in a fake phishing email. Table 7.14 presents the results of other components of the framework for PhishGuru intervention.

7.2.5 Challenges in administering real-world phishing studies

We have taken measures in this study to address many lessons learned from earlier work. Real-world studies can provide more ecological validity and richer data than laboratory studies, but are often difficult to conduct. The challenges we faced included making sure the study emails reached participants' inboxes, maintaining participants' privacy, avoiding contamination between study conditions, and coordinating with relevant third parties.

Simulated emails may get deleted before they reach the user's inbox if, for instance, filters determine that the message is spam. Additionally, since many web-browsers come equipped with anti-phishing tools, one has to be careful that the study material isn't blocked. In particular, one should be aware of the possibility that study websites might end up on a black-list. To be prepared for problems of this nature, we registered multiple dummy domains and prepared multiple sets of emails as backup. Furthermore, since email reading behavior may be different over university holidays than it is during the regular semester, we carefully timed the study schedule so that the study emails were not sent during university holidays.

In order to maintain the privacy of the participants, study administrators should not/cannot col-

Table 7.14: Human-in-the-loop analysis. We used the framework developed by Cranor to analyze the PhishGuru intervention [50]. We found that people are able to process the information presented by PhishGuru and apply that information to identify future phishing emails.

Component		**Outcome**
Communication		Training, active, this may be the best moment for the communication (teachable moment)
Communication impediments	Environmental stimuli	Users just want to get to the website
	Interference	It is difficult for organizations to send fake phishing emails to their customers
Personal variables	Demographics and personal characteristics	The system should work for everyone
	Knowledge and experience	The system should work for everyone
Intentions	Attitudes and beliefs	Users like the PhishGuru concept and the intervention (as seen in the feedback from the studies)
	Motivation	Since the intervention is presented at the teachable moment, users read the intervention and not fall for phishing attacks in the future
Capabilities		Not beyond the capability of users
Communication delivery	Attention switch	Since it is active training, users notice the communication
	Attention maintenance	Users read the PhishGuru intervention completely, reasons: it is fun, story based, and presents actionable items. From walkthroughs and focus group studies, we found that users of all age read the complete intervention
Communication processing	Comprehension	From the data collected, participants who saw the training are less likely to fall for phishing attacks in future. Thus they appear to comprehend it. However, we haven't specifically tested comprehension
	Knowledge acquisition	All studies show that people were able to process and extract knowledge from PhishGuru intervention
Application	Knowledge retention	All studies show that people retain the concepts and procedures from PhishGuru intervention and make better decisions in identifying the emails
	Knowledge transfer	All studies show that people transfer the concepts and procedures from PhishGuru intervention to other situations and make better decisions in identifying the emails. However, we only tested near transfer
Behavior		From both real-world and laboratory studies, it is clear that PhishGuru training results in behavior change

lect any personal information. Furthermore, to understand the users' behavior over time, users' responses must be tracked in a way that respects their privacy. We accomplished this in the study by assigning each participant an anonymous hash and using only that hash to track them.

To avoid subject contamination, study designers should try to minimize the chance that participants in different conditions will interact with each other; such interactions may invalidate the study data. Working to prevent these interactions, study designers must ensure that the study sample is embedded within a large, geographically separate population. In the previous field study, significant contamination occurred because study participants all worked on one floor of an office building. In the current study, even though all participants were from the same university campus, they represented a small fraction of the campus population and were spread across 26 departments and many buildings, which limited contamination.

It is important to coordinate with any relevant third parties that might be affected by the study. We worked very closely with ISO in both the design and implementation stages of this study. In addition, ISO helped us get permission from the Institutional Review Board (IRB), coordinate with campus help desks, and get permission from all campus offices spoofed in the study. As a courtesy and to minimize accidental external interference in the study, researchers should work with system administrators and help desk officials of the organization to inform them about the study. If possible, researchers should also provide system administrators with a "canned" response they can use to respond to any inquiries from participants. This helps minimize the chance that system administrators will send an email to the entire population warning them to avoid opening an email that is actually part of the study (we have seen this happen in a prior study). Finally, it is essential that any university phishing study go through the university's IRB. Having a well-defined plan to address the challenges mentioned here can help prevent potential difficulties in the review process.

7.2.6 Discussion

In this section, we investigated the effectiveness of PhishGuru, an embedded training methodology that teaches people about phishing during their normal use of email. We showed that, even 28 days after training, users trained by PhishGuru were less likely to click on a link in a simulated phishing email than those who were not trained. Furthermore, users who saw the training intervention twice were less likely to give information to fake phishing websites than those who only saw the training intervention once. Additionally, results from this study indicate that training users to recognize phishing emails using PhishGuru does not increase their concern towards email in general or cause them to make more false positive mistakes. Another surprising result was that around 90% of the participants who eventually clicked on the link in an email did so within 8 hours of the time the email was sent. We believe this behavior extends to other university populations, though non-university populations may behave quite differently when reading emails. A demographic analysis showed that young people (in the 18-25 age group) were more likely than older participants to consistently fall for phishing emails on all days of the study. This suggests a need for: (1) training before college; and (2) training that specifically targets high school and college students.

The study presented in this section addresses some of the limitations of earlier laboratory (Chapter 6) and real-world (Section 7.1) studies of PhishGuru. To address these limitations, we employed a larger sample size, extended the study duration, counterbalanced the email and training interventions, minimized the chance of contamination from participants talking about the study amongst themselves, and provided good incentives for participants to complete the post-study survey. In the process of addressing these limitations, we successfully showed that PhishGuru can be deployed both on a large scale and in the real world as an embedded training system that educates users about phishing during their regular use of email. This study included only a small fraction of the campus population due to IRB requirements that participants opt in to the study before receiving any study emails. However, if this deployment had been done as a real training exercise—that is, without an academic IRB requirement—we believe it would have been easy to train the entire campus with only minimal changes to the study setup.

This study affirms prior research suggesting that the PhishGuru methodology is an unobtrusive way to train users about phishing. Some comments from the post-study survey include: (1) "I really liked the idea of sending CMU students fake phishing emails and then saying to them, essentially, HEY! You could've just gotten scammed! You should be more careful – here's how...." (2) "I think the idea of using something fun, like a cartoon, to teach people about a serious subject is awesome!" (3) "Pictures and short examples are the best way for me not to ignore these kinds of messages."

Furthermore, the fact that knowledge gained from the training materials is retained for at least 28 days suggests that very frequent interventions, which could annoy users, are not necessary. In practice, this should be balanced with the fact that repeated training does improve user performance; a proper trade-off between usability and accuracy can and should be optimized.

In addition to increasing user awareness about phishing emails, there was evidence that the study had the unintended consequence of assessing both the users' awareness of proper response channels for phishing attacks and the ability of the ISO to react to phishing attacks. Many users properly contacted the ISO help desk to alert them of the emails, either by phone or through the official email address. However, some were apparently unaware of the ISO's role in protecting the campus, and instead contacted some other "trusted source" like a professor or departmental system administrator to seek advice. This suggests that the ISO may want to explore ways to increase awareness of the proper channels for reporting phishing attacks and other cyber security related issues. In a real deployment of PhishGuru, training interventions could be one way to distribute this information to the public.

This study, along with the study discussed in Section 7.1, is proof that it is possible to effectively educate users about security in the real world and on a large scale. Findings from this study suggest that security researchers and practitioners should implement user training as a complementary strategy to other technological solutions for security problems.

Chapter 8

Other Implementations of Phishing Education

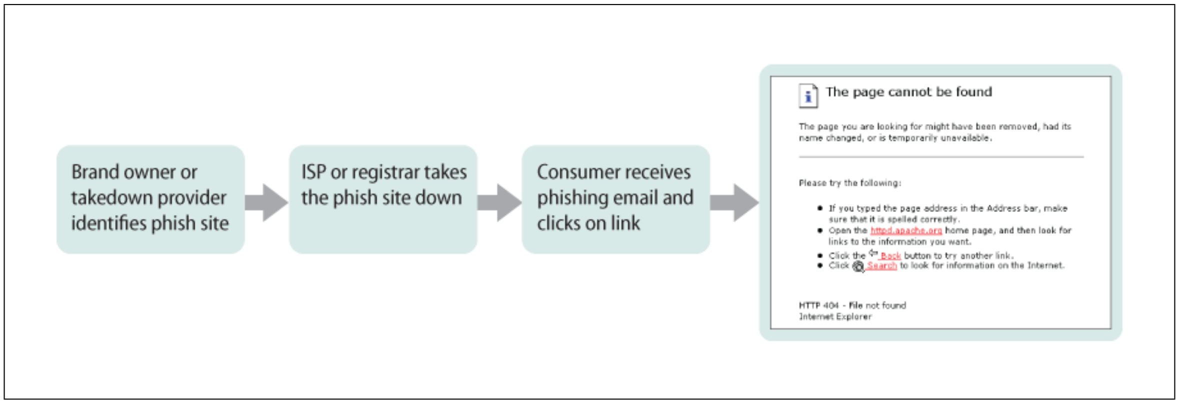
In addition to PhishGuru, we have also developed two other approaches to phishing education: (1) the Anti-Phishing Working Group (APWG) landing page, where a training intervention is presented to users who go to a phishing website that has already been taken down; and (2) Anti-Phishing Phil, an online game that teaches people how to identify phishing URLs. We discuss the details of the APWG landing page in Section 8.1 and the evaluation of Phil in Section 8.2.

8.1 Anti-Phishing Working Group landing page

In Section 8.1.1, we discuss the concept of the landing page and how it evolved from the PhishGuru results. In Section 8.1.2, we present the infrastructure we developed to collect data on people accessing the landing page. In Section 8.1.3, we discuss the design evolution of the landing page and results from the focus group studies we conducted to design the landing page. In Section 8.1.4, we present results from the data we collected. In Section 8.1.5, we discuss some implications of the results.

8.1.1 Landing page concept

Most phishing websites are taken down sooner or later. Brand owners, takedown vendors, or law enforcement get the phishing sites taken down. Figure 8.1 illustrates a common scenario; when users click on a phishing link in an email that takes them to a website that has already been taken down, they are directed to a 404 error or told that "the page cannot be found." At the 2007 APWG eCrime Researchers Summit, researchers and industry representatives discussed how the


Brand owner or takedown provider identifies phish site
ISP or registrar takes the phish site down
Consumer receives phishing email and clicks on link

The page cannot be found

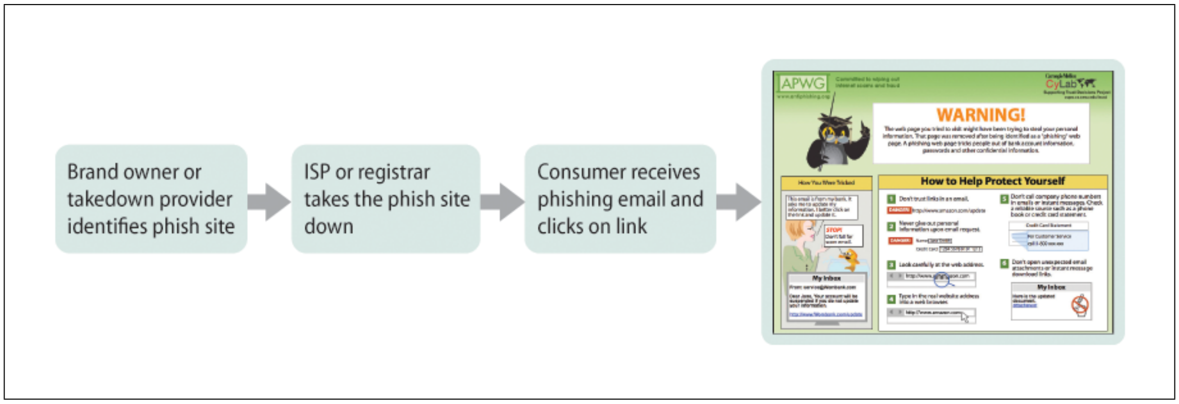
The page you are looking for might have been removed, had its name changed, or is temporarily unavailable.

Please try the following:

- If you typed the page address in the Address bar, make sure that it is spelled correctly.
- Open the httpd.apache.org home page, and then look for links to the information you want.
- Click the Back button to try another link.
- Click Search to look for information on the Internet.

HTTP 404 - File not found
Internet Explorer

Figure 8.1: The current situation. Users are presented with "The page cannot be found" message when they click on a link to a phish site that has been taken down.


Brand owner or takedown provider identifies phish site
ISP or registrar takes the phish site down
Consumer receives phishing email and clicks on link
APWG
CyLab
How You Were Tricked
STOP!
Don't fall for scam email.
My Inbox

WARNING!

The web page you tried to visit might have been trying to steal your personal information. That page was removed after being identified as a 'phishing' web page. A phishing web page tricks people out of bank account information, passwords and other confidential information.

How to Help Protect Yourself

1 Don't trust links in an email.
2 Never give out personal information upon email request.
3 Look carefully at the web address.
4 Type in the real website address into a web browser.
5 Don't call company phone numbers in emails or instant messages. Check a reliable source such as a phone book or credit card statement.
6 Don't open unexpected email attachments or instant message download links.

Figure 8.2: APWG landing page. Users are presented with a version of the PhishGuru intervention when they click on a link to a phish site that has been taken down.

results of the study presented in Section 6.2 could be applied to this scenario. A Bank of America representative mentioned that they were also working internally to present a warning page to users who clicked on a link to a site that had already been taken down [22].

Following the Summit, we started working with the APWG-IPC (Internet Policy Committee) to apply the PhishGuru results to the creation of a solution the industry could use. We designed a landing page that used the PhishGuru intervention instead of the 404 error message. Figure 8.2 shows how the new landing page appears to users. We started working on both the infrastructure and content of the intervention.

8.1.2 Infrastructure

In order to make this an industry-wide initiative which any organization could use, a publicly available sub-domain was set up on the APWG website – http://education.apwg.org/. Information about the project was posted on this website. The English version of the landing page was hosted at http://education.apwg.org/r/en/. Since this page was going to be translated into many other languages, it was decided that users would be redirected to a specific language depending on the default language of their web browser. As of March 29, 2009, people had volunteered to translate the landing page into Arabic, Bulgarian, Catalan, Danish, Dutch, French, German, Hebrew, Japanese, Korean, Romanian, Spanish, and Swedish. The French landing page is available at http://education.apwg.org/r/fr/.

The success of this implementation depended on brands adopting the landing page as their redirect page. To that end, we created a "how to" file that provided information about redirecting sites to the landing page. In particular, the document included information about redirecting in Apache and Microsoft Internet Information Services (IIS). We suggested that, while doing the redirect, the ISP or registrar should add the URL in the URL request to the landing page. This is achieved by adding the phishing URL after a "?" in the HTTP request to the landing page. This redirect can be done in Apache and IIS in the following ways:

- Apache
  - Create a .htaccess file in the directory where the phishing site was stored. Note the leading dot on the .htaccess filename.
  - The .htaccess file should contain the text:
Redirect 301 /the-phishing-page.html http://education.apwg.org/r/en? www.phishsite.com/the-phishing-page.html
  - In the above text, "the-phishing-page.html" should be replaced with the filename of the phishing webpage that was taken down. "www.phishsite.com/the-phishing-page.html" should be replaced by the full URL of the phish site that was taken down. Note that there are two things that need to be replaced by the full URL of the phish site. For example, "the-phishing-page.html" could be "signin.html" and "www.phishsite.com/the-phishing-page.html" could be "yourcompany.com/update/signin.html"
- IIS
  - Change the HttpRedirect property for the resource to:
http://education.apwg.org/r/en?the-phishing-page.html, PERMANENT
  - Note that "the-phishing-page.html" should be replaced with the filename of the phishing webpage that was taken down. For example, "the-phishing-page.html" could be "signin.html."

Since we have access to the log files of the landing page, we can create a list of phishing URLs. Using RSync, the APWG webserver for http://education.apwg.org/ replicates the logs onto the

CUPS (CyLab Usable Privacy and Security) laboratory server once a week. If registrars and ISPs implement the changes discussed above, the log entries will capture the link users click before they are redirected to the landing page.

Table 8.1 includes one sample log entry. The number 74.276.172.102 represents the IP address of the client which made the request for the landing page. Next is the time [03/Oct/2008:01:11:57 -0500] the server finished processing the request. After that is the request line, which sits between the quotes; this includes the method used by the client (in this example GET), the request URL (/r/en?mail.millenniumantenna.com/icons/image.htm/), and the protocol the client used (HTTP/1.1). The entry then provides the status code the server sends back to the client. Codes beginning with 2 are successful responses, while codes beginning with 4 are errors. Next, the entry lists the site the client was referred from; in this example, it is mail.yahoo.com. Finally, the entry lists the user-agent HTTP request header, which contains identifying information the client browser reports about itself.

Table 8.1: Sample of the APWG landing page log entry.

```
74.276.172.102
[03/Oct/2008:01:11:57 -0500]
“GET /r/en?mail.millenniumantenna.com/icons/image.htm/ HTTP/1.1”
200 283
“http://us.mg2.mail.yahoo.com/dc/blank.html?bn=1096.40&.intl=us”
“Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.8.1.17)
Gecko/20080829 Firefox/2.0.0.17”
```

The CUPS server also receives APWG’s feed of reported phishing emails (emails sent to reportphishing@antiphishing.org). We correlated this data with the log data to find out which emails led most users to visit the landing page. The email feed from APWG contains the entire email with all of the headers. This data gave insight into the most vulnerable emails; that is, those emails that contained links leading users to phishing websites.

8.1.3 Design evolution and evaluation

In this section, we discuss intervention design challenges, design decisions, design iterations, and two focus group studies conducted to revise the interventions.

There were some challenges in developing the landing page content. One challenge, which we had also faced when designing PhishGuru, was to limit the content to one browser page. This makes it so users don’t have to scroll to read the instructions. When designing the landing page, we also had to consider the fact that users access it from hand held devices. This means that the content has to be really light weight, with as few images as possible. We worked to find a compromise between these challenges and the factors needed to produce a useful final product.

The APWG Internet Policy Committee provided suggestions on what instructions needed to be in the intervention. The first design the committee suggested is provided in Figure 8.3 and Figure 8.4. This design was two pages long. We reviewed the design and suggested some changes in the intervention (Figure 8.5 presents the revised version of the intervention). The main things we removed were instructions not relevant to phishing and some information on available resources.

Focus group studies

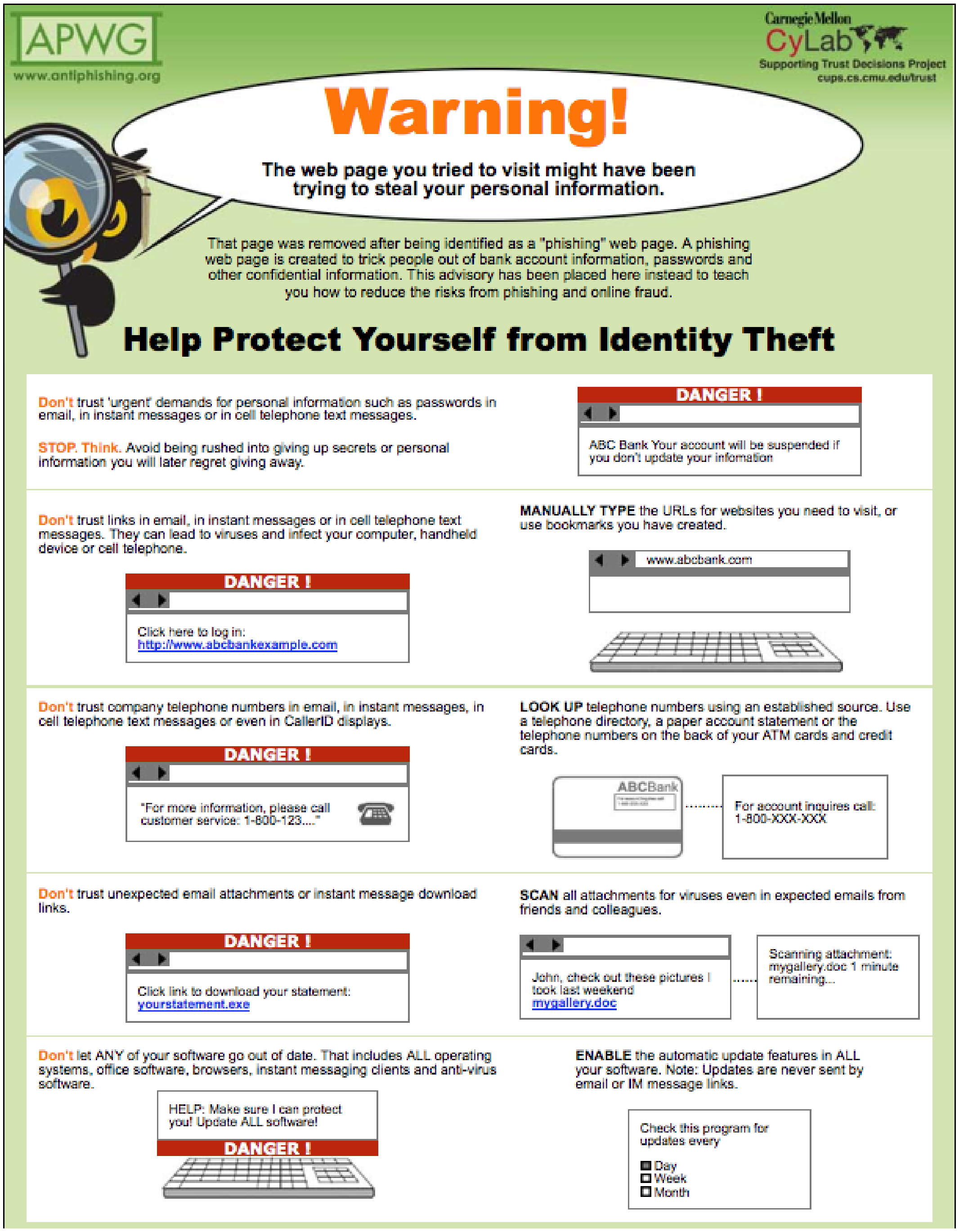
We conducted two focus group studies to evaluate the effectiveness of the content in the intervention. In this section, we discuss each study's setup and results. We explain how the results guided us in developing an effective landing page.

Focus group study I: The first focus group we conducted was a 2-hour session at Carnegie Mellon University with nine participants. There were 5 females and 4 males. The average age of participants was 26 years (min: 18, max: 53). Participants received an average of 20 emails per day (min: 5, max: 35). None of the participants knew what phishing was. Participants came from a variety of backgrounds – business, arts and science, social work, fine arts, nursing, music, and psychology. Two of the participants had only a high school degree. Using a wall projector, we began the focus group by demonstrating how someone might click on a link in a phishing email and arrive at the landing page. We then showed them what they might see on a landing page. We discussed details of 3 versions of the intervention: (1) the committee draft (Figure 8.3 and Figure 8.4); (2) a condensed draft (Figure 8.5); and (3) PhishGuru (Figure 8.6). We provided participants with a color printout of the designs and gave them pencils so they could provide feedback on the printouts. We also voice recorded the entire session.

1. **Committee draft**: Participants felt that the committee draft was too much to read. Most participants said they would not read past the "Help Protect Yourself" headline because there were too many things for them to parse and understand. Due to the length of the text, six of the nine participants said they would only read until the first instruction. Two said they would read the entire intervention, while only one was willing to read the additional resources.

   Participants had difficulty navigating through the intervention (i.e. participants read down the left column and then down the right rather than reading across). Participants were confused by the browser images; some participants thought they were text entry forms, while others had trouble understanding the DANGER! text above the images.

   All of the participants wanted information about what to do after reading the landing page. Since this page warns users about links in emails or instant messages, participants were



APWG
www.antiphishing.org

Carnegie Mellon
CyLab
Supporting Trust Decisions Project
cups.cs.cmu.edu/trust

Warning!

The web page you tried to visit might have been trying to steal your personal information.

That page was removed after being identified as a "phishing" web page. A phishing web page is created to trick people out of bank account information, passwords and other confidential information. This advisory has been placed here instead to teach you how to reduce the risks from phishing and online fraud.

Help Protect Yourself from Identity Theft

Don't trust 'urgent' demands for personal information such as passwords in email, in instant messages or in cell telephone text messages.

STOP. Think. Avoid being rushed into giving up secrets or personal information you will later regret giving away.

DANGER !

ABC Bank Your account will be suspended if you don't update your information

Don't trust links in email, in instant messages or in cell telephone text messages. They can lead to viruses and infect your computer, handheld device or cell telephone.

DANGER !

Click here to log in:
http://www.abcbankexample.com

MANUALLY TYPE the URLs for websites you need to visit, or use bookmarks you have created.

www.abcbank.com

Don't trust company telephone numbers in email, in instant messages, in cell telephone text messages or even in CallerID displays.

DANGER !

"For more information, please call customer service: 1-800-123...."

LOOK UP telephone numbers using an established source. Use a telephone directory, a paper account statement or the telephone numbers on the back of your ATM cards and credit cards.

ABCBank

For account inquires call:
1-800-XXX-XXX

Don't trust unexpected email attachments or instant message download links.

DANGER !

Click link to download your statement:
yourstatement.exe

SCAN all attachments for viruses even in expected emails from friends and colleagues.

John, check out these pictures I took last weekend
mygallery.doc

Scanning attachment:
mygallery.doc 1 minute remaining...

Don't let ANY of your software go out of date. That includes ALL operating systems, office software, browsers, instant messaging clients and anti-virus software.

HELP: Make sure I can protect you! Update ALL software!

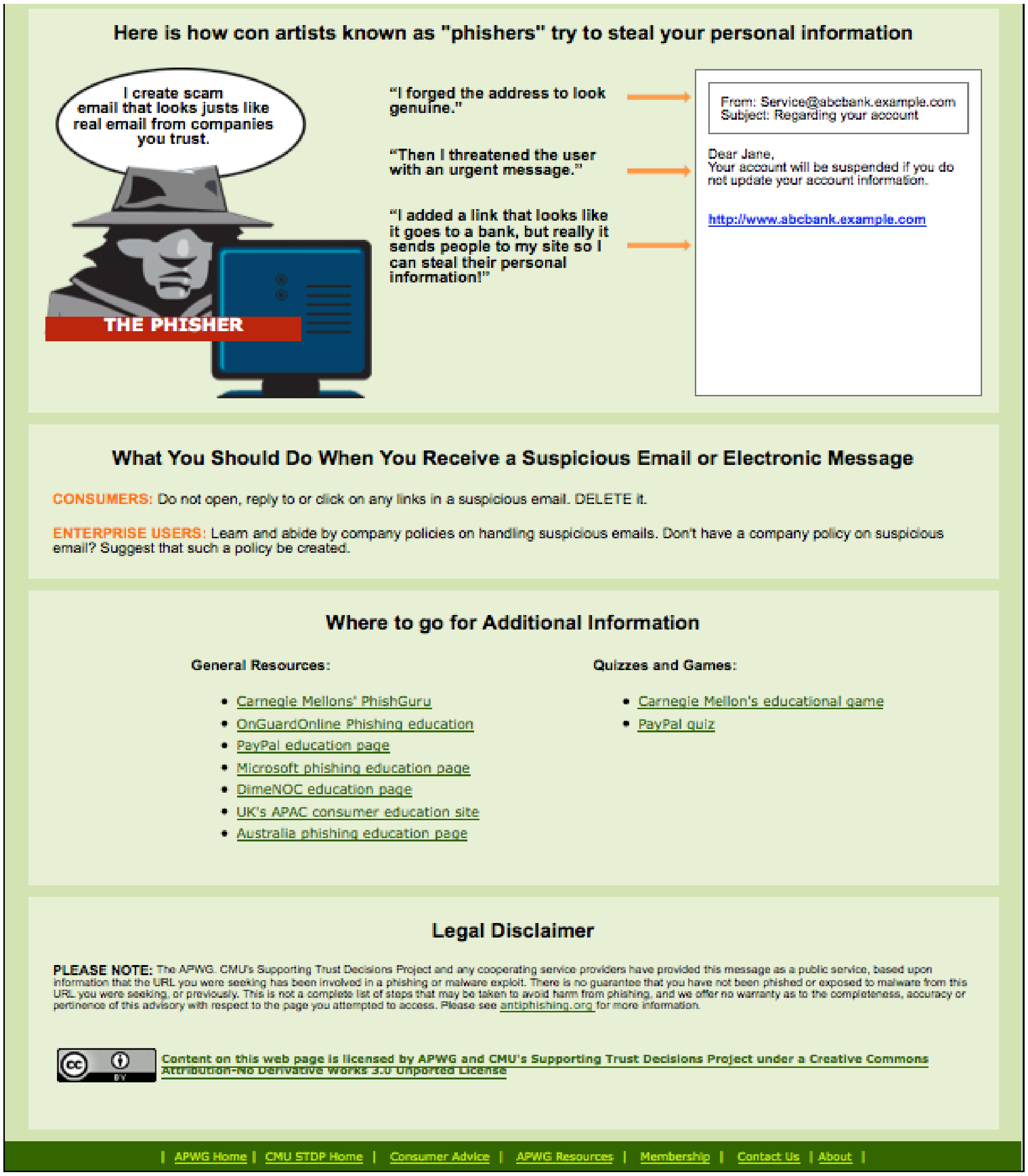
DANGER !

ENABLE the automatic update features in ALL your software. Note: Updates are never sent by email or IM message links.

Check this program for updates every

■ Day
□ Week
□ Month

Figure 8.3: IPC committee version page 1. This design was created by the IPC at APWG. This has information about phishing but also information about software updates and viruses. Participants in the study felt that this design was too long. They also had difficulty navigating through the intervention.



I create scam email that looks justs like real email from companies you trust.
THE PHISHER

Here is how con artists known as "phishers" try to steal your personal information

"I forged the address to look genuine."

"Then I threatened the user with an urgent message."

"I added a link that looks like it goes to a bank, but really it sends people to my site so I can steal their personal information!"

From: Service@abcbank.example.com
Subject: Regarding your account

Dear Jane,
Your account will be suspended if you do not update your account information.

http://www.abcbank.example.com

What You Should Do When You Receive a Suspicious Email or Electronic Message

CONSUMERS: Do not open, reply to or click on any links in a suspicious email. DELETE it.

ENTERPRISE USERS: Learn and abide by company policies on handling suspicious emails. Don't have a company policy on suspicious email? Suggest that such a policy be created.

Where to go for Additional Information

General Resources:

- Carnegie Mellons' PhishGuru
- OnGuardOnline Phishing education
- PayPal education page
- Microsoft phishing education page
- DimeNOC education page
- UK's APAC consumer education site
- Australia phishing education page

Quizzes and Games:

- Carnegie Mellon's educational game
- PayPal quiz

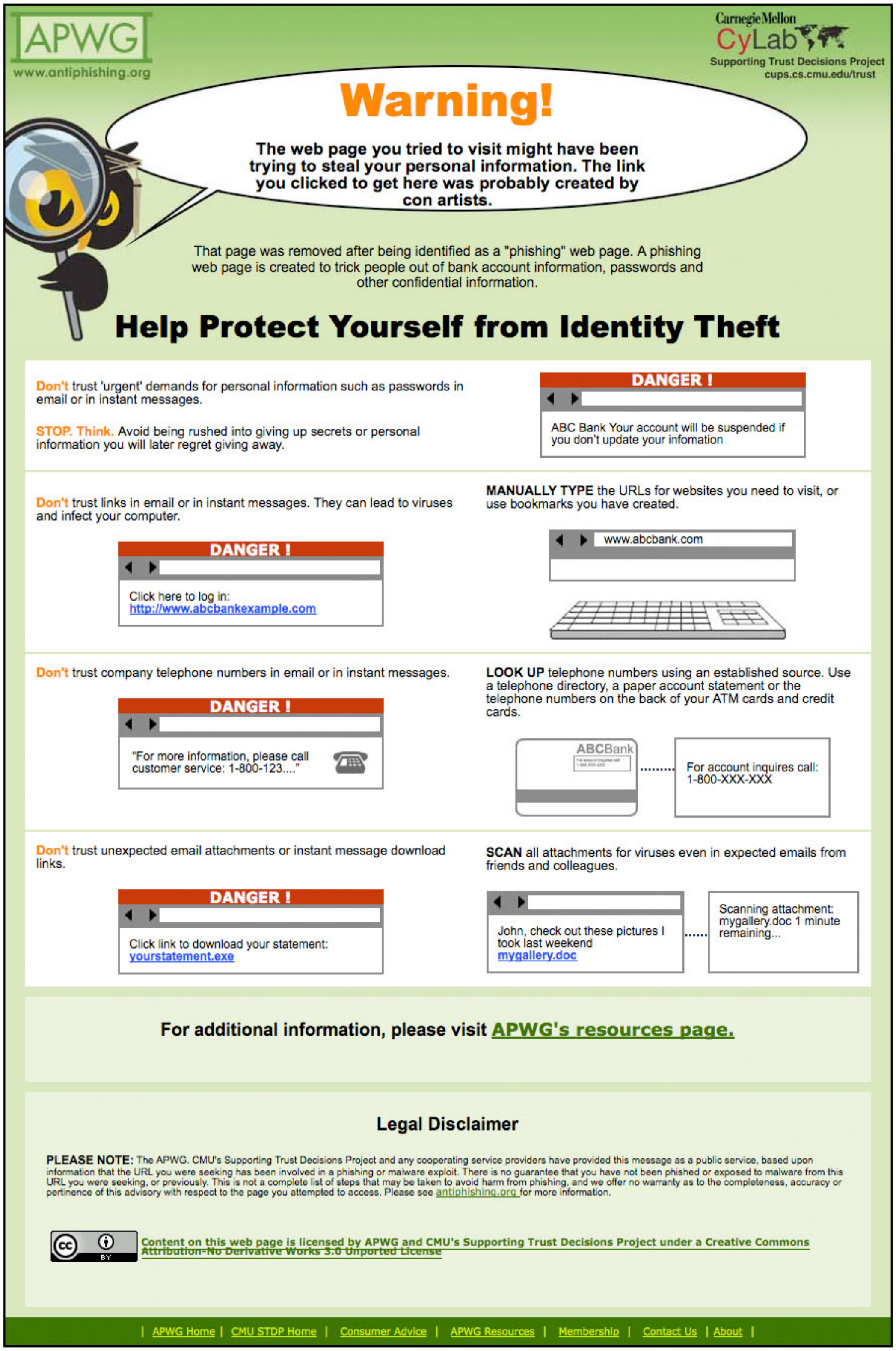
Legal Disclaimer

PLEASE NOTE: The APWG. CMU's Supporting Trust Decisions Project and any cooperating service providers have provided this message as a public service, based upon information that the URL you were seeking has been involved in a phishing or malware exploit. There is no guarantee that you have not been phished or exposed to malware from this URL you were seeking, or previously. This is not a complete list of steps that may be taken to avoid harm from phishing, and we offer no warranty as to the completeness, accuracy or pertinence of this advisory with respect to the page you attempted to access. Please see antiphishing.org for more information.

Content on this web page is licensed by APWG and CMU's Supporting Trust Decisions Project under a Creative Commons Attribution-No Derivative Works 3.0 Unported License

| APWG Home | CMU STDP Home | Consumer Advice | APWG Resources | Membership | Contact Us | About |

Figure 8.4: IPC committee version page 2. Participants in the focus group studies liked the phisher character, but did not understand the meaning of "Enterprise Users." Participants also did not like the idea of links to other sources.



APWG
www.antiphishing.org

Carnegie Mellon CyLab
Supporting Trust Decisions Project
cups.cs.cmu.edu/trust

Warning!

The web page you tried to visit might have been trying to steal your personal information. The link you clicked to get here was probably created by con artists.

That page was removed after being identified as a "phishing" web page. A phishing web page is created to trick people out of bank account information, passwords and other confidential information.

Help Protect Yourself from Identity Theft

Don't trust 'urgent' demands for personal information such as passwords in email or in instant messages.

STOP. Think. Avoid being rushed into giving up secrets or personal information you will later regret giving away.

DANGER !
ABC Bank Your account will be suspended if you don't update your infomation

Don't trust links in email or in instant messages. They can lead to viruses and infect your computer.

DANGER !
Click here to log in:
http://www.abcbankexample.com

MANUALLY TYPE the URLs for websites you need to visit, or use bookmarks you have created.

www.abcbank.com

Don't trust company telephone numbers in email or in instant messages.

DANGER !
"For more information, please call customer service: 1-800-123...."

LOOK UP telephone numbers using an established source. Use a telephone directory, a paper account statement or the telephone numbers on the back of your ATM cards and credit cards.

ABCBank

For account inquires call:
1-800-XXX-XXX

Don't trust unexpected email attachments or instant message download links.

DANGER !
Click link to download your statement:
yourstatement.exe

SCAN all attachments for viruses even in expected emails from friends and colleagues.

John, check out these pictures I took last weekend
mygallery.doc

Scanning attachment: mygallery.doc 1 minute remaining...

For additional information, please visit APWG's resources page.

Legal Disclaimer

PLEASE NOTE: The APWG. CMU's Supporting Trust Decisions Project and any cooperating service providers have provided this message as a public service, based upon information that the URL you were seeking has been involved in a phishing or malware exploit. There is no guarantee that you have not been phished or exposed to malware from this URL you were seeking, or previously. This is not a complete list of steps that may be taken to avoid harm from phishing, and we offer no warranty as to the completeness, accuracy or pertinence of this advisory with respect to the page you attempted to access. Please see antiphishing.org for more information.

Content on this web page is licensed by APWG and CMU's Supporting Trust Decisions Project under a Creative Commons Attribution-No Derivative Works 3.0 Unported License

| APWG Home | CMU STDP Home | Consumer Advice | APWG Resources | Membership | Contact Us | About |

Figure 8.5: This is the revised version of Figure 8.3 and 8.4. To make it short, we removed the phisher image, links to other sources, and a few instructions. Participants in the studies liked that it was short, but wished the phisher was in the design.

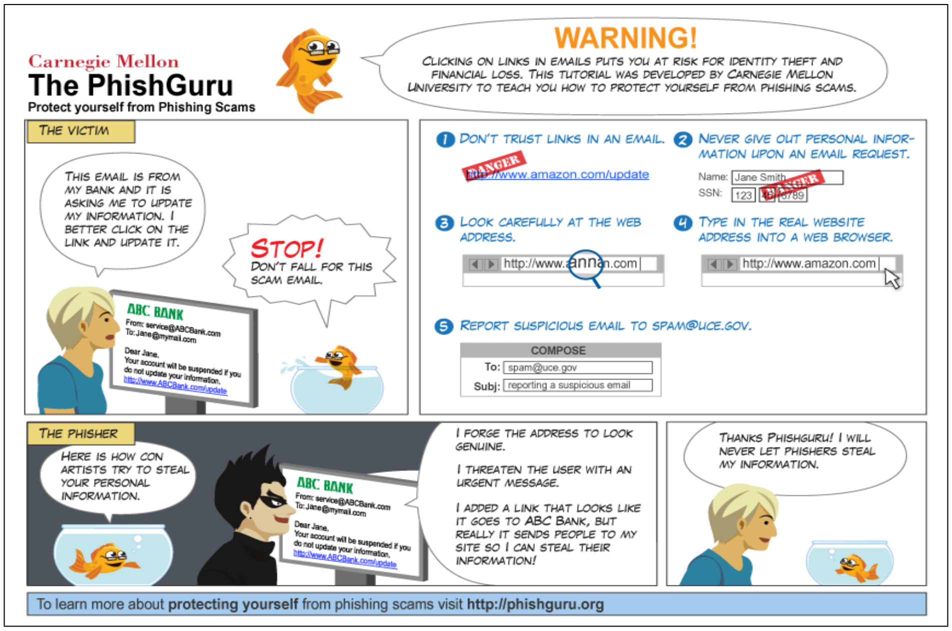

Carnegie Mellon
The PhishGuru
Protect yourself from Phishing Scams
WARNING!
CLICKING ON LINKS IN EMAILS PUTS YOU AT RISK FOR IDENTITY THEFT AND FINANCIAL LOSS. THIS TUTORIAL WAS DEVELOPED BY CARNEGIE MELLON UNIVERSITY TO TEACH YOU HOW TO PROTECT YOURSELF FROM PHISHING SCAMS.
THE VICTIM
THIS EMAIL IS FROM MY BANK AND IT IS ASKING ME TO UPDATE MY INFORMATION. I BETTER CLICK ON THE LINK AND UPDATE IT.
STOP! DON'T FALL FOR THIS SCAM EMAIL.
ABC BANK
From: service@ABCBank.com
To: Jane@mymail.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.ABCBank.com/update
1 DON'T TRUST LINKS IN AN EMAIL.
DANGER
http://www.amazon.com/update
2 NEVER GIVE OUT PERSONAL INFORMATION UPON AN EMAIL REQUEST.
Name: Jane Smith
SSN: 123 45 6789
DANGER
3 LOOK CAREFULLY AT THE WEB ADDRESS.
http://www.annan.com
4 TYPE IN THE REAL WEBSITE ADDRESS INTO A WEB BROWSER.
http://www.amazon.com
5 REPORT SUSPICIOUS EMAIL TO SPAM@UCE.GOV.
COMPOSE
To: spam@uce.gov
Subj: reporting a suspicious email
THE PHISHER
HERE IS HOW CON ARTISTS TRY TO STEAL YOUR PERSONAL INFORMATION.
ABC BANK
From: service@ABCBank.com
To: Jane@mymail.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.ABCBank.com/update
I FORGE THE ADDRESS TO LOOK GENUINE.
I THREATEN THE USER WITH AN URGENT MESSAGE.
I ADDED A LINK THAT LOOKS LIKE IT GOES TO ABC BANK, BUT REALLY IT SENDS PEOPLE TO MY SITE SO I CAN STEAL THEIR INFORMATION!
THANKS PHISHGURU! I WILL NEVER LET PHISHERS STEAL MY INFORMATION.
To learn more about protecting yourself from phishing scams visit http://phishguru.org

Figure 8.6: One version of the PhishGuru intervention. We used a standard comic strip font. Participants in the studies did not like the font or the phisher character. Most of them said the phisher looked like Batman.

concerned about clicking links on this page to get more information.

Participants liked some of the visual features of the page, particularly the owl. Participants thought the phisher character was very appropriate and wanted to see him at the top of the page.

2. **Condensed version**: Participants felt that even though this intervention (see Figure 8.5) was short, it was also too long to read it completely. Only three of the nine participants said they would read the entire intervention, while the rest said they would only read until the first instruction.

   Participants recognized the same problems in this version as in the committee version. Since navigation in this page was the same as in the committee version, participants felt it was not easy to navigate through the instructions. Participants were also confused by the browser images.

   Participants noted that the phisher was removed in this intervention. Participants suggested bringing the phisher back and putting him at the top. One participant mentioned "Put the scary guy back."

3. **PhishGuru**: Most participants said they would be much more likely to read the PhishGuru version (see Figure 8.6) of the intervention completely. Participants found it both entertaining and informative, and they liked the PhishGuru gold fish character. One participant said, "I

> really enjoy it and I would probably read it because it is entertaining, but people wouldn't take it seriously." One participant exclaimed: "Exactly what we were looking for" after looking at the PhishGuru version. Some had concerns about the perceived credibility of a comic strip. Participants also raised concerns that the comic book font was hard to read and didn't look very official. All participants liked the fish character; some said, "I like this gold fish" and "It is just cute." All of the participants agreed that they would definitely read the PhishGuru intervention, adding that the comic script would appeal to their parents and grandparents. One participant mentioned "I think my grandma would get the comic spot on." Participants mentioned that "Guru" sounds very official, as it refers to some knowledgeable person.
>
> Participants stated that having "Carnegie Mellon" in the intervention added credibility to the presented information. This sentiment could have been more prevalent because all of the participants were from Pittsburgh.
>
> Although participants liked the PhishGuru gold fish character, they complained that the phisher character looked like batman. They mentioned that the phisher character was not evil enough or scary.

To test whether it really would appeal to older people, we conducted a second focus group study, which is discussed below.

Focus group study II: The second focus group we conducted was a 2.5 hour session with six participants at The Jewish Community Center of Greater Pittsburgh. We worked with AgeWell's Independent Adult Services Department to recruit participants who were more than 65 years old. This study involved 3 females and 3 males. The average age of the participants was 76 years (min: 66, max: 83), with one participant declining to give her age. Participants received an average of 7.3 emails per day (min: 2, max: 15). None of the participants knew what phishing was. Participants had a variety of educational backgrounds – business, english, architecture, medical, and engineering. One participant had a high school degree. As with the first focus group study, we began this focus group by demonstrating how someone might click on a link in a phishing email and arrive at the landing page; we then showed the group what they might see on the landing page. None of the participants knew how easy it is to spoof an email address and send fake emails pretending to come from legitimate organizations.

Using feedback from the first focus group study, we revised the condensed and PhishGuru versions of the landing page. In the revised condensed version, the instructions were made exactly the same as the PhishGuru version, but the rest was kept the way it was in the earlier condensed version. In the PhishGuru version, we changed the comic font to Helvetica and cleaned up some text. In focus group study II, we discussed details of 3 versions of the intervention: (1) the committee draft (Figure 8.3 and Figure 8.4); (2) the revised condensed draft (Figure 8.7); and (3) PhishGuru (Figure 8.8). We provided color printouts of these designs to participants and gave them pencils so they could provide feedback on the printouts. We also voice recorded the entire session.

1. **Committee draft**: Participants in this study, like in the first study, responded negatively to the committee draft. Most of the participants said they would not read the complete page. Since the page was long, most of the participants mentioned that they would only scan the whole intervention, while two said they would read it completely.

   Participants in this study also were confused by the browser images. Some participants thought "http://www.abcbankexample.com" in second instruction (see Figure 8.3) was a link to click.

   Participants had mixed reactions about the characters in this intervention. Almost all participants liked the phisher character, saying "He looks like a thief or a criminal." Most did not like the owl character. Some were confused by the owl's magnifying glass and mortar board hat; others thought the owl's hand was a duck's head.

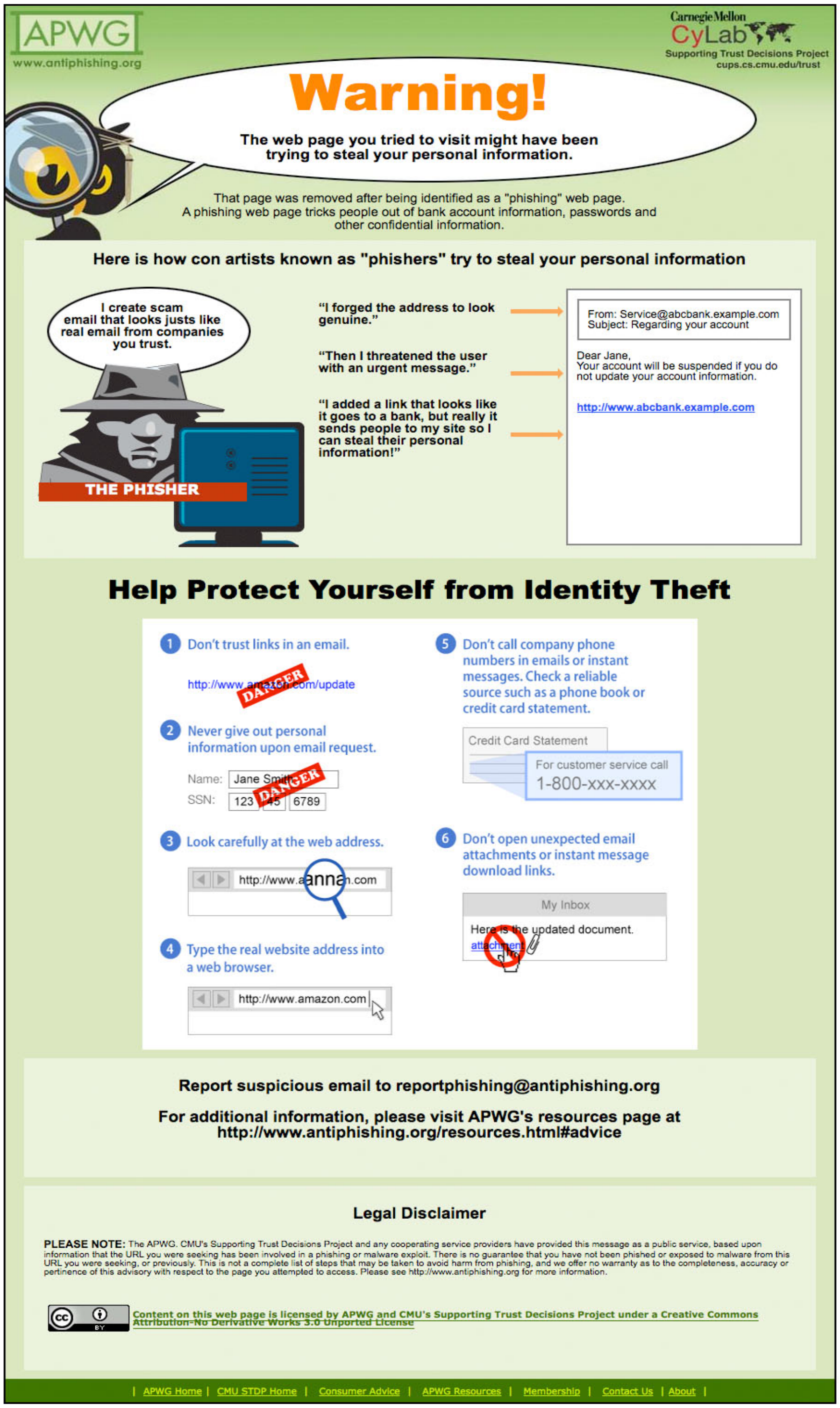
2. **Revised condensed version**: Participants liked the fact that the revised condensed version (see Figure 8.7) was short and had less text. Some participants mentioned even though it is shorter than committee version, it is still long and therefore would not read it completely.

   Participants enjoyed the images in the instructions in this version. Participants liked the emphasis on things with the "DANGER" symbol; they said that this would get their attention and get them to read it. All participants liked the fact that the instructions had pictures they could understand. One participant said "This is more pictorial than the first one ... so much better."

   As in the previous focus group study, almost all participants liked having the Carnegie Mellon name in the intervention. This could again be due to the fact that all of the participants were from Pittsburgh.

   One instruction generated a lot of discussion, most of the participants did not know that calling a phone number in a phishing message could be dangerous. One participant mentioned "I didn't know that – even if you call a company phone number you will get into trouble."

3. **PhishGuru**: Participants were attracted to the PhishGuru intervention, stating that it was fun to read and that people of all ages would read it. Participants were interested in the cartoon format and characters. All participants liked the fish character. Some reactions from the participants about the interventions were: "I like this one ... I really do," "eye catchy," and "1 [the committee version] & 2 [the condensed version] are business like and 3 is fun." All participants said they would read the complete intervention. All participants agreed that having characters is good and likely to attract readers' attention. Some participants did not think the phisher character looked evil enough, preferring the phisher in the revised



APWG
www.antiphishing.org
Carnegie Mellon CyLab
Supporting Trust Decisions Project
cups.cs.cmu.edu/trust

Warning!

The web page you tried to visit might have been trying to steal your personal information.

That page was removed after being identified as a "phishing" web page.
A phishing web page tricks people out of bank account information, passwords and other confidential information.

Here is how con artists known as "phishers" try to steal your personal information

I create scam email that looks justs like real email from companies you trust.
THE PHISHER

"I forged the address to look genuine."

"Then I threatened the user with an urgent message."

"I added a link that looks like it goes to a bank, but really it sends people to my site so I can steal their personal information!"

From: Service@abcbank.example.com
Subject: Regarding your account

Dear Jane,
Your account will be suspended if you do not update your account information.

http://www.abcbank.example.com

Help Protect Yourself from Identity Theft

1 Don't trust links in an email.

2 Never give out personal information upon email request.

3 Look carefully at the web address.

4 Type the real website address into a web browser.

5 Don't call company phone numbers in emails or instant messages. Check a reliable source such as a phone book or credit card statement.

6 Don't open unexpected email attachments or instant message download links.

Report suspicious email to reportphishing@antiphishing.org

For additional information, please visit APWG's resources page at http://www.antiphishing.org/resources.html#advice

Legal Disclaimer

PLEASE NOTE: The APWG. CMU's Supporting Trust Decisions Project and any cooperating service providers have provided this message as a public service, based upon information that the URL you were seeking has been involved in a phishing or malware exploit. There is no guarantee that you have not been phished or exposed to malware from this URL you were seeking, or previously. This is not a complete list of steps that may be taken to avoid harm from phishing, and we offer no warranty as to the completeness, accuracy or pertinence of this advisory with respect to the page you attempted to access. Please see http://www.antiphishing.org for more information.

Content on this web page is licensed by APWG and CMU's Supporting Trust Decisions Project under a Creative Commons Attribution-No Derivative Works 3.0 Unported License

| APWG Home | CMU STDP Home | Consumer Advice | APWG Resources | Membership | Contact Us | About |

Figure 8.7: Second revised version of the landing page. We made the instructions look exactly the same as in the PhishGuru design (Figure 8.6). We made the phisher more prominent in this design and added the email address where any complaints or reports could be sent.

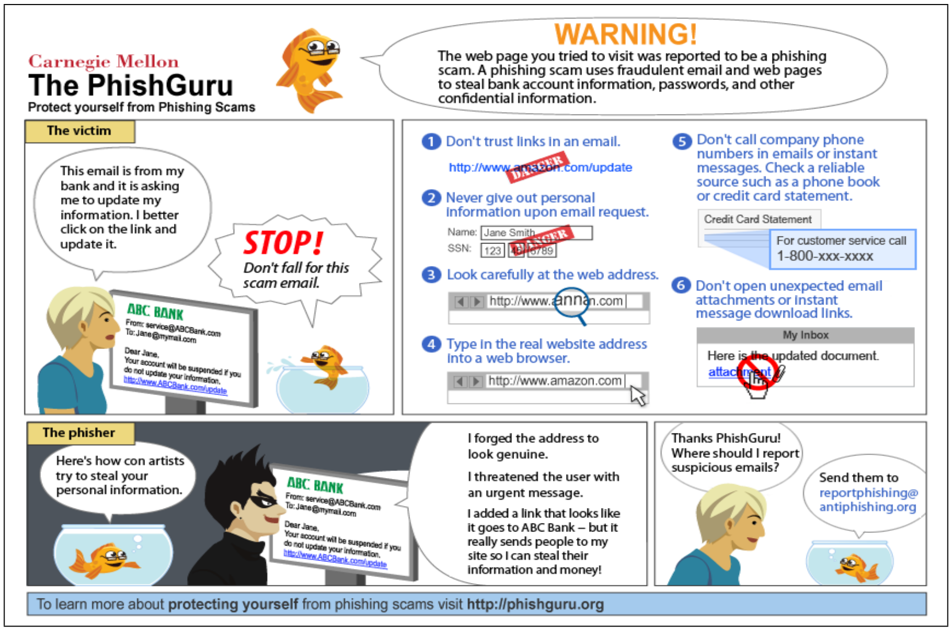
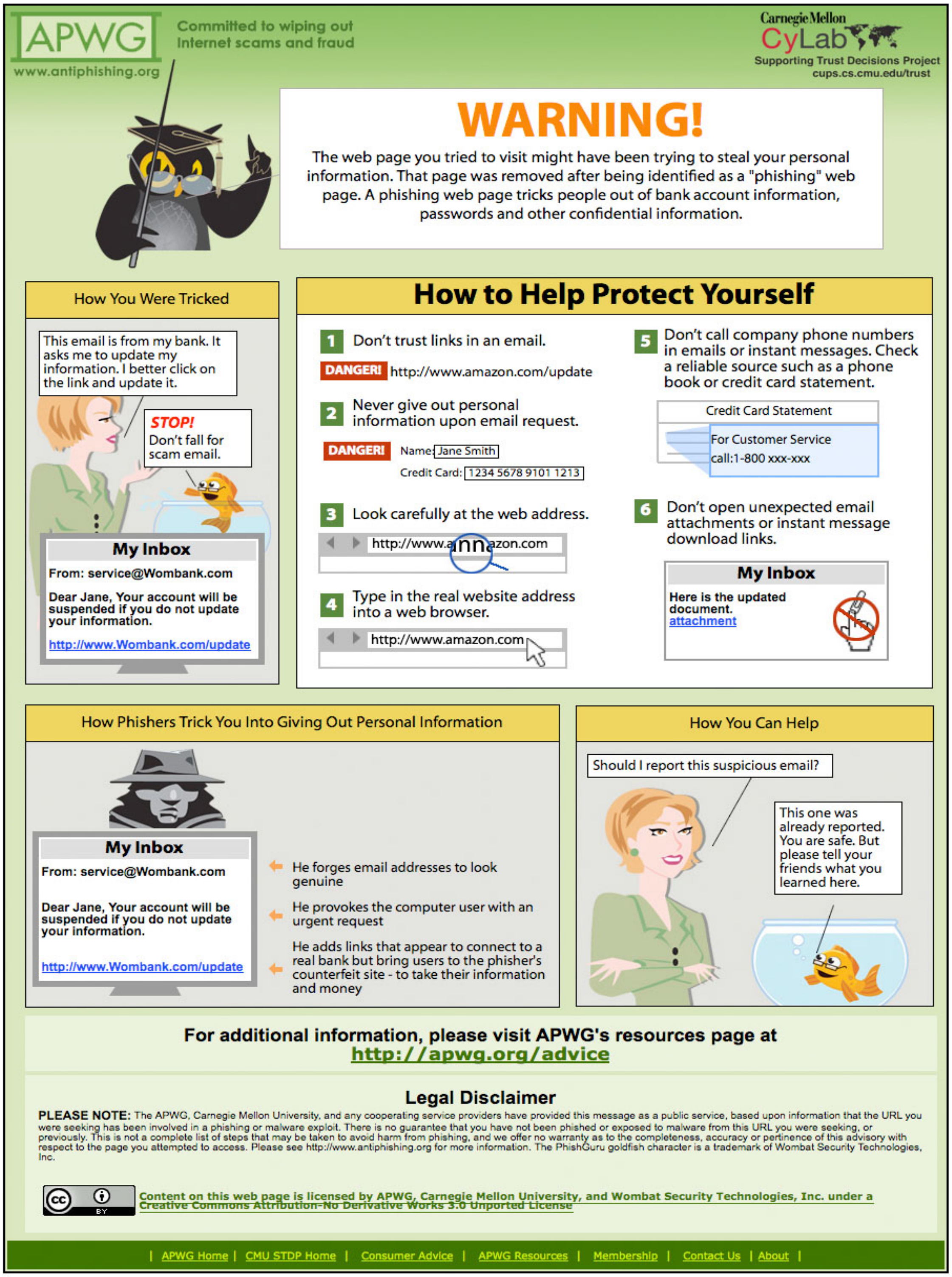

Carnegie Mellon
The PhishGuru
Protect yourself from Phishing Scams
WARNING!
The web page you tried to visit was reported to be a phishing scam. A phishing scam uses fraudulent email and web pages to steal bank account information, passwords, and other confidential information.
The victim
This email is from my bank and it is asking me to update my information. I better click on the link and update it.
STOP! Don't fall for this scam email.
ABC BANK
From: service@ABCBank.com
To: Jane@mymail.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.ABCBank.com/update
1 Don't trust links in an email.
http://www.amazon.com/update
DANGER
2 Never give out personal information upon email request.
Name: Jane Smith
SSN: 123 45 6789
DANGER
3 Look carefully at the web address.
http://www.annazon.com
4 Type in the real website address into a web browser.
http://www.amazon.com
5 Don't call company phone numbers in emails or instant messages. Check a reliable source such as a phone book or credit card statement.
Credit Card Statement
For customer service call 1-800-xxx-xxxx
6 Don't open unexpected email attachments or instant message download links.
My Inbox
Here is the updated document.
attachment
The phisher
Here's how con artists try to steal your personal information.
ABC BANK
From: service@ABCBank.com
To: Jane@mymail.com
Dear Jane,
Your account will be suspended if you do not update your information.
http://www.ABCBank.com/update
I forged the address to look genuine.
I threatened the user with an urgent message.
I added a link that looks like it goes to ABC Bank – but it really sends people to my site so I can steal their information and money!
Thanks PhishGuru! Where should I report suspicious emails?
Send them to reportphishing@antiphishing.org
To learn more about protecting yourself from phishing scams visit http://phishguru.org

Figure 8.8: PhishGuru revised version. This is a revised version of Figure 8.6, with new fonts and cleaned-up text. Participants in the studies enjoyed reading this version and also suggested that all age groups would read this intervention in its entirety.

committee version (Figure 8.7). No participants had concerns about people not taking the revised PhishGuru intervention seriously.

Overall, these focus group studies showed that people in both younger and older age groups like the PhishGuru intervention and would be likely to read it. The main reasons people liked PhishGuru was its character style, pictorial representation, use of narrative, and comic format. Using the results of the focus group studies, the IPC at APWG was convinced to make an intervention for the landing page that is more like PhishGuru.

We used the focus group results to start developing the intervention for the landing page. One constraint was that, in the real world, people might access the landing page from a variety of devices, such as desktop PCs, hand held devices like PDAs, or mobile phones. This basically meant that the entire intervention needed to be in plain html, with as few images as possible. This would make loading of the page easier for all types of devices. Based on the feedback from the focus group studies and these constraints, we developed the intervention shown in Figure 8.9. This is the version available to users as of Jan 16, 2009.



APWG

Committed to wiping out Internet scams and fraud

www.antiphishing.org

Carnegie Mellon CyLab

Supporting Trust Decisions Project
cups.cs.cmu.edu/trust

WARNING!

The web page you tried to visit might have been trying to steal your personal information. That page was removed after being identified as a "phishing" web page. A phishing web page tricks people out of bank account information, passwords and other confidential information.

How You Were Tricked

This email is from my bank. It asks me to update my information. I better click on the link and update it.

STOP! Don't fall for scam email.

My Inbox

From: service@Wombank.com

Dear Jane, Your account will be suspended if you do not update your information.

http://www.Wombank.com/update

How to Help Protect Yourself

1 Don't trust links in an email.

DANGER! http://www.amazon.com/update

2 Never give out personal information upon email request.

DANGER! Name: Jane Smith
Credit Card: 1234 5678 9101 1213

3 Look carefully at the web address.

http://www.amnazon.com

4 Type in the real website address into a web browser.

http://www.amazon.com

5 Don't call company phone numbers in emails or instant messages. Check a reliable source such as a phone book or credit card statement.

Credit Card Statement
For Customer Service
call:1-800 xxx-xxx

6 Don't open unexpected email attachments or instant message download links.

My Inbox

Here is the updated document.
attachment

How Phishers Trick You Into Giving Out Personal Information

My Inbox

From: service@Wombank.com

Dear Jane, Your account will be suspended if you do not update your information.

http://www.Wombank.com/update

He forges email addresses to look genuine

He provokes the computer user with an urgent request

He adds links that appear to connect to a real bank but bring users to the phisher's counterfeit site - to take their information and money

How You Can Help

Should I report this suspicious email?

This one was already reported. You are safe. But please tell your friends what you learned here.

For additional information, please visit APWG's resources page at http://apwg.org/advice

Legal Disclaimer

PLEASE NOTE: The APWG, Carnegie Mellon University, and any cooperating service providers have provided this message as a public service, based upon information that the URL you were seeking has been involved in a phishing or malware exploit. There is no guarantee that you have not been phished or exposed to malware from this URL you were seeking, or previously. This is not a complete list of steps that may be taken to avoid harm from phishing, and we offer no warranty as to the completeness, accuracy or pertinence of this advisory with respect to the page you attempted to access. Please see http://www.antiphishing.org for more information. The PhishGuru goldfish character is a trademark of Wombat Security Technologies, Inc.

Content on this web page is licensed by APWG, Carnegie Mellon University, and Wombat Security Technologies, Inc. under a Creative Commons Attribution-No Derivative Works 3.0 Unported License

| APWG Home | CMU STDP Home | Consumer Advice | APWG Resources | Membership | Contact Us | About |

Figure 8.9: Final version of the landing page. This is the final version available online at http://education.apwg.org/r/en/. We are analyzing the logs for this page.

8.1.4 Results

In this section, we discuss some of the analyses performed with the data from the logs and email feeds. All data we received from APWG was ported to MySQL; analyses were done using MySQL statements, perl scripting, and R 1.24 on Mac OS X. Not all data from the logs was directly usable for analysis. We filtered out entries using the logic presented in Table 8.2 to get unique URLs from the logs. We used these unique URLs to get all other data discussed in this section.

In this section, we present: (1) the complete analysis of the logs we collected and (2) results of the feature analysis performed on the emails retrieved from the email feed, which were done using the URLs in the logs.

Aggregate view of the data

To analyze the results from the logs, we used the pseudocode presented in Table 8.2. The idea was to use only log entries that contained '/r/en/?', as these entries were created because users clicked on links in emails to websites that had been taken down. We removed entries which contained the terms 'ORIGINAL_PHISH_URL' or 'www.phishsite.com' or 'the-phishing-page.html.' These are involved in the documentation on how to implement the landing page; therefore, these may be hits by organizations or vendors testing the landing page. The data we used for this analysis was from Oct 1, 2008 to March 1, 2009. After filtering the entries, we viewed three segments of the data: (1) the whole—to see the total number of hits the landing page was getting; (2) only those URLs with more than 5 hits; and (3) only those URLs with less than or equal to 5 hits. We analyzed the data in different approaches (e.g. looking at the frequency distribution of hits corresponding to the URLs, getting the IP ranges/subnets from take down vendors and organizations and removing them from logs). We found that there was a significant jump in hits after 5 as compared to less than or equal to 5. We experimented with varying the cutoff, but found that 5 point mark provided us with reliable data to analyze end-users viewing the landing page against organizations or take down vendors testing the page. We also vetted this approach with a couple of take down vendors that we interact with. We also confirmed that the IPs that are in greater than 5 set did not have large number of hits. This verifies that the data in greater than 5 set does not have hits from organizations or take down vendors but people actually viewing the landing page. We believe that URLs with less than or equal to 5 hits are mostly takedown vendors or organizations testing their implementation of the landing page or checking whether the landing page is active. The organizations and takedown vendors we worked with have said anecdotally that they check the phished URLs for the redirect at least a couple times. We believe that URLs with greater than 5 hits are real users clicking on links that have been taken down. One can argue that this list may include some hits from organizations or takedown vendors, but this will be hard to detect empirically with our current methods. Table 8.2 also presents the number of entries at every given stage of the pseudocode.

This shows that many hits are not relevant to the data analysis (i.e. about 46,283 hits are removed between step 2 and step 7).

Table 8.2: Pseudocode for getting unique URLs from the log entries. Values presented are for Oct 1, 2008 to March 1, 2009. The entries column presents the log entries available until that time.

Step	Pseudocode for getting unique URLs from the log entries	Entries
1	Push the log entries into MySQL database	2,489,667
2	Extract entries which have '/r/en?' in the request URL	109,005
3	Remove entries with 'ORIGINAL_PHISH_URL' in the request URL	96,217
4	Remove entries with 'www.phishsite.com' in the request URL	64,162
5	Remove entries with 'the-phishing-page.html' in the request URL	63,729
6	Remove entries that have information only like 'http:/' or 'section=SiteKey&amp' in the request URL	62,722
7	Analyze the entire data set for different statistics	62,722
8	Analyze URLs which has less than or equal to 5 hits for different statistics	5,973
9	Analyze URLs which has greater than 5 hits for different statistics	56,699

We believe the landing page has created many teachable moments in which users have been trained to avoid falling for future phishing attacks. Table 8.3 presents statistics for the hits on the landing page. From the entire data, there were 62,722 total hits on the page; among these hits, there were 3,763 unique URLs. These statistics suggest that at least 56,699 "teachable moments" have been created using the landing page.

Table 8.4 shows statistics for how long people are clicking on these URLs in emails. Column 4 shows that people click on links an average of 34.9 days from the first time the URL appeared in the logs. Researchers and organizations should develop tools to help lower this average and protect users from phishing emails.

Using the IP addresses from the log entries, we identified the country of origin for users viewing the landing page. We saw that most hits (87.8%) came from the United States (see Table 8.5). This may be due to the fact that, at least for the time being, the brands who have adopted the landing page are mainly from the US. This also may be because the organizations being phished are mostly from the US [15]. This result may change as more brands around the world start using the landing page. We also found that around 98% of the total hits on the landing page were from the top 10 countries on the list.

Table 8.3: Comprehensive view of the APWG landing page logs for the period Oct 1, 2008 to March 1, 2009.

Statistics	Whole data set	Less than or equal to 5 hits	Greater than 5 hits
Number of unique URLs	3,763	3,639	124
Total Hits for all unique URLs	62,722	6,023	56,699
Maximum number of hits for a single URL	3,875	5	3,875
Minimum number of hits for a single URL	1	1	6
Average number of hits per URL	16.7	1.7	457.25
Median number of hits for the URLs	2	2	149
Standard deviation for the URLs	158.6	0.6	752.9

Table 8.4: Days between the first time a URL appears and the last time it appears for the period Oct 1, 2008 to March 1, 2009. Values presented in parentheses are in minutes.

Statistics	Whole data	Less than or equal to 5 hits	Greater than 5 hits
Maximum number of days	145.8	73.7	145.8
Minimum number of days	0	0	0.01 (19)
Average number of days	6.4	5.5	34.9
Median number of days	0	0	25.5
Standard deviation	16.9	15.0	35.7

Email feature analysis

To study the emails that correspond to the URLs we retrieved from landing page logs, we compared the unique URLs from the logs to the URLs in the APWG email feed. We retrieved emails with the unique URL (from the logs) that were embedded in the emails. Using all of the data from the logs and the email feed from Oct 1, 2008 to March 1, 2009, we found 67 URL matches. We manually went through the 67 emails and analyzed the features in the emails. Around 95% of the emails were from Bank of America; the rest were from other popular financial institutions and goverment agencies.

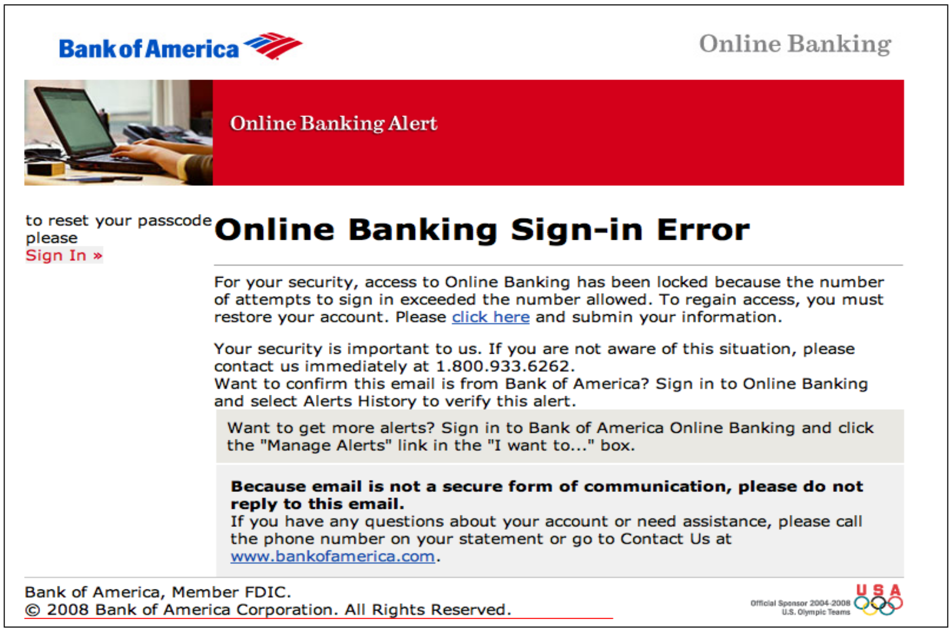
Most of the emails had features similar to legitimate emails. Ninety-one percentage of the emails

Table 8.5: Percentage of hits from the top 10 countries. Analysis was performed on the entire data set.

S.No.	**URL**	**Percentage of hits**
1.	United States	87.8
2.	United Kingdom	5.9
3.	Japan	1.5
4.	Canada	1.4
5.	Israel	1.0
6.	Hong Kong	0.6
7.	Brazil	0.6
8.	Netherlands	0.5
9.	Australia	0.5
10.	Europe Union	0.4

had some form of logo or banner at the top of the email. As Dhamija et al. showed [53], the fact that these logos and banners look legitimate is one of the main reasons why people fall for phishing emails. Seventy-three percentage of the emails had some sort of footer with logos; in particular, Bank of America emails had an Olympics logo in the bottom right corner (See Figure 8.10). In some cases phishers used an exact replica of the legitimate emails. Figure 8.10 presents both the legitimate and the phishing email (found in the email feed) for the same scenario–"Online Banking Sign-in Error" for Bank of America.

Most of the emails provide compelling scenarios why people should click on the link in the emails. Seventy-seven percentage of the emails had some form of urgent actionable message in their subject line (e.g. "Online Banking Alert - Your Online Banking Account is Locked" and "Your Account Has been Temporarily Suspended"). Most of the emails (85%) asked users to click on the link and update or verify their account information. Only a few of the emails presented a scenario in which users were told that they had a new message in their "secure message" inbox and that they should click on the enclosed link to view the message. Most of the emails requested account information, but some explicitly asked recipients to provide "your username or SSN and your password." Many scenarios were presented in these emails; one of the common ones was "We recently have determined that different computers have logged in your Bank of America Online Banking account, and multiple password failures were present before the logons." Most of the emails mentioned some form of consequence (e.g. account suspension), and 13% of the emails suggested that there would be consequences if the recipient failed to act within a given time frame. Almost all of these emails mentioned a deadline of 2 days or 48 hours from the time the email was sent. One common message regarding the timeline was "Please update your records on or before 48 hours, a failure to update your records will result in a temporary hold on your funds."



Bank of America

Online Banking

Online Banking Alert

to reset your passcode please
Sign In »

Online Banking Sign-in Error

For your security, access to Online Banking has been locked because the number of attempts to sign in exceeded the number allowed. To regain access, you must restore your account. Please click here and submin your information.

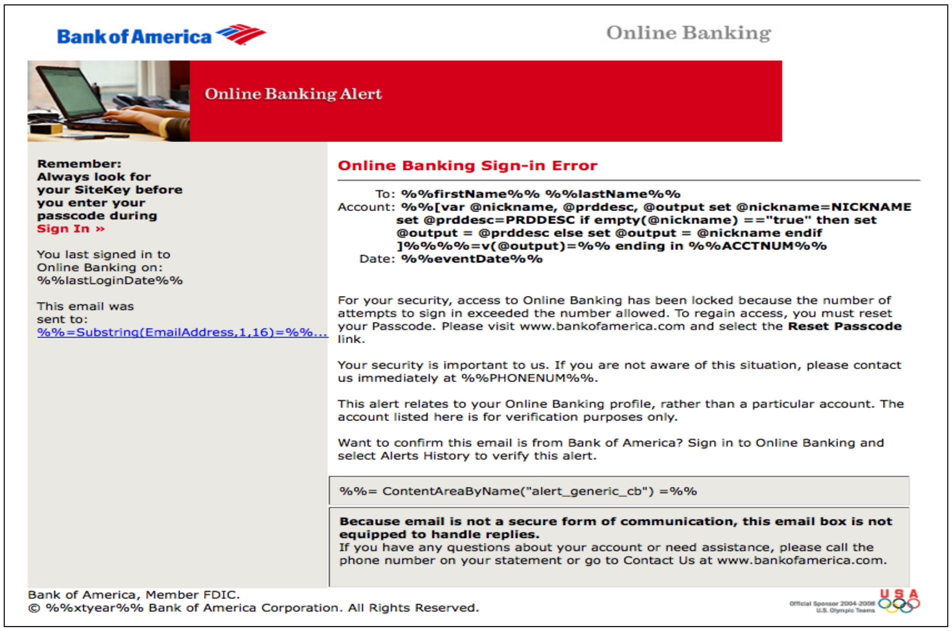
Your security is important to us. If you are not aware of this situation, please contact us immediately at 1.800.933.6262.
Want to confirm this email is from Bank of America? Sign in to Online Banking and select Alerts History to verify this alert.

Want to get more alerts? Sign in to Bank of America Online Banking and click the "Manage Alerts" link in the "I want to..." box.

Because email is not a secure form of communication, please do not reply to this email.
If you have any questions about your account or need assistance, please call the phone number on your statement or go to Contact Us at www.bankofamerica.com.

Bank of America, Member FDIC.
© 2008 Bank of America Corporation. All Rights Reserved.

Official Sponsor 2004-2008 U.S. Olympic Teams USA



Bank of America

Online Banking

Online Banking Alert

Remember:
Always look for your SiteKey before you enter your passcode during
Sign In »

You last signed in to Online Banking on:
%%lastLoginDate%%

This email was sent to:
%%=Substring(EmailAddress,1,16)=%%...

Online Banking Sign-in Error

```
     To: %%firstName%% %%lastName%%
Account: %%[var @nickname, @prddesc, @output set @nickname=NICKNAME
         set @prddesc=PRDDESC if empty(@nickname) =="true" then set
         @output = @prddesc else set @output = @nickname endif
         ]%%%%=v(@output)=%% ending in %%ACCTNUM%%
   Date: %%eventDate%%
```

For your security, access to Online Banking has been locked because the number of attempts to sign in exceeded the number allowed. To regain access, you must reset your Passcode. Please visit www.bankofamerica.com and select the **Reset Passcode** link.

Your security is important to us. If you are not aware of this situation, please contact us immediately at %%PHONENUM%%.

This alert relates to your Online Banking profile, rather than a particular account. The account listed here is for verification purposes only.

Want to confirm this email is from Bank of America? Sign in to Online Banking and select Alerts History to verify this alert.

%%= ContentAreaByName("alert_generic_cb") =%%

Because email is not a secure form of communication, this email box is not equipped to handle replies.
If you have any questions about your account or need assistance, please call the phone number on your statement or go to Contact Us at www.bankofamerica.com.

Bank of America, Member FDIC.
© %%xtyear%% Bank of America Corporation. All Rights Reserved.

Official Sponsor 2004-2008 U.S. Olympic Teams USA

Figure 8.10: Top: Phishing email from the APWG email dump that pretends to come from Bank of America. Bottom: A real email from Bank of America to their customers. All information with "%" are used to customize the emails with personal information.

We found that the emails contained many formatting and grammatical errors. Some errors in the emails were: "If this is not completed by octobre 03, 2008," "If we do no receive," and "check you account profile." Another error was that one email said "please supply all of the following information," but it offered no list of what information the recipient should provide. One email was entirely center aligned to the page; this email also presented many telephone numbers in a table format. Some numbers were 1-800 and some were non 1-800 numbers. Some emails contained text in an entirely bold font, some had text that was all blue, and some contained a combination of black, blue, and red text. Again, results from Chapter 4 show that non-experts do use font color as a signal to make their decision.

We found that phishers are still using traditional techniques to con people. Some of the domain names used for sending these phishing emails look similar to legitimate ones (e.g. onlinebanking@alert.bank0famerica.com, where the 'o' in 'of' is replaced with '0'). Around 76% of the emails have text like "Click here to continue" or "Signin" or "click here" as a link in the email. These sentences are linked to the phishing websites. The rest of the emails had some sort of disguised link leading to the phishing website.

We also found that, in some emails, there was a mismatch between the subject line and the content of the email. For example, in one email, the subject line was "Online Banking Alert," but the email content scenario was "Online Banking Sign-in Error." In another email, the subject line was "online Banking Sign-in Error," but the content of the email was about verifying account information. There were also a mismatch between brands in the sender address and the content in the email. For example, the content of one email was for Bank of America, while the From address was from a different well-known financial institution.

Ninety-six percentage of the emails were not customized for the recipient with any form of personal information. Three of the emails included some form of personal information: (1) customer ID, (2) account type and ending number, and (3) account type. It is not clear whether this was really customized for the recipient or not.

To increase the chances of people falling for these attacks, phishers are also using other techniques to con people. Twenty-one percent of the emails invited readers to call for clarification or assistance. A typical example was "If you are not aware of this situation, please contact us immediately at 1.800.123.456." As expected, most of those numbers don't match real numbers. We also found that the phone numbers were different in many of the emails. In its legitimate emails, Bank of America uses different phone numbers depending on the location of the customer or the nature of the email. It looks like, as in other respects, phishers are emulating what legitimate organizations are doing.

8.1.5 Discussion

In this section, we discussed a real-world implementation of PhishGuru. In general, we found that a majority of the phishing emails we found from the APWG feed are using the same phishing kits to generate these emails. Since most phishing emails replicate legitimate emails, researchers and industry could reap substantial benefits by creating a corpus of legitimate emails, studying their features, and incorporating these features into email filters. Phishing emails haven't changed much over time, remaining relatively unsophisticated and containing a great number of errors in grammar and formatting. Most of the emails in the log analysis contained information relating to the account details, asking users to click on a link in the email to update their account details.

The results of this analysis confirm that the instructions in the PhishGuru and the landing page cover features contained in most phishing emails. Users who know these cues will be better able to identify phishing emails and avoid being victims of phishing. In particular: (1) we found that all emails had disguised links; this relates to the PhishGuru instruction – "Don't trust links in an email." (2) We found that most of the emails ask for account details; this relates to the PhishGuru instruction – "Never give out personal information upon email request." (3) Most URLs in the emails look similar to legitimate ones; this relates to the PhishGuru instruction – "Look carefully at the web address." (4) Some emails lure people into calling a fake number; this relates to the PhishGuru instruction – "Don't call company phone numbers in emails or instant messages."

8.2 Anti-Phishing Phil

> This section is largely a reproduction of a paper co-authored with Steve Sheng, Alessandro Acquisti, Lorrie Cranor, and Jason Hong and accepted at TOIT [111]. An earlier version of the paper was co-authored with Steve Sheng, Bryant Magnien, Alessandro Acquisti, Lorrie Cranor, Jason Hong, and Elizabeth Nunge and published at SOUPS 2007 [181].

As another implementation of phishing training, we used learning science principles to develop Anti-Phishing Phil,[1] an educational game. Phil was designed to train users about phishing attacks, motivating them to learn by embedding training into a fun activity. The highly interactive nature of the game allows it to teach users to distinguish legitimate links from fraudulent ones; it also provides users with immediate opportunities to practice this procedure multiple times. Anti-Phishing Phil complements PhishGuru by providing an entertaining platform for the rapid repetition and feedback needed to teach more difficult anti-phishing procedures. Phil is currently being commercialized by

[1] http://cups.cs.cmu.edu/antiphishing_phil/

Wombat Security Technologies.[2] In this section of the chapter, we will discuss the design and evaluation of Phil. In Section 8.2.1, we present the design of Anti-Phishing Phil and describe the ways in which we applied instructional design principles to the design of the game. In Section 8.2.2, we present a laboratory study evaluation. In Section 8.2.3, we present results from a field study.

8.2.1 Design of Anti-Phishing Phil

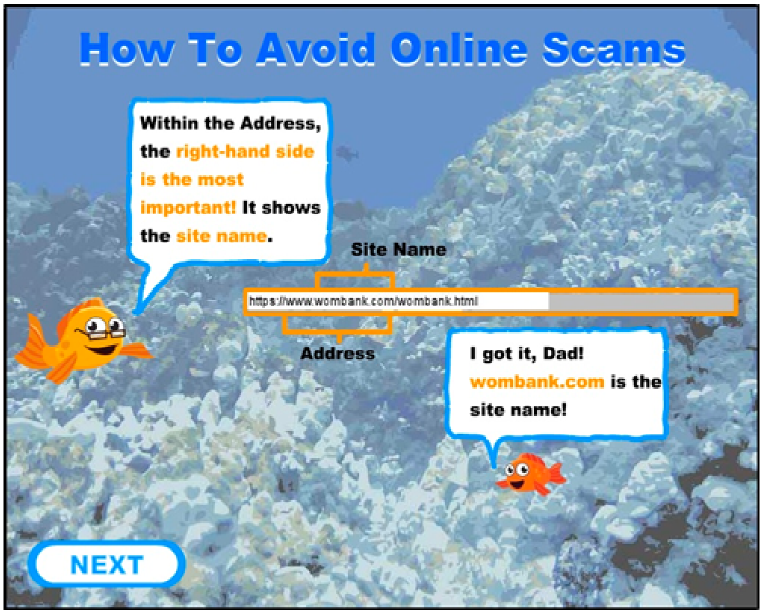
The main character of the game is a young fish named Phil. Phil wants to eat worms so he can grow up to be a big fish, but has to be careful of phishers who try to trick him with fake worms (which represent phishing attacks). Each worm is associated with a URL, and Phil's job is to eat all of the real worms (which have URLs of legitimate websites) and reject all of the bait (which have phishing URLs) before running out of time. The other character is PhishGuru, who is an experienced fish. He helps Phil out by providing tips on how to identify fake worms (and hence, phishing websites).

The game is split into four rounds, each two minutes long. Before each round begins, users view a short tutorial that provides anti-phishing tips, as shown in Figure 8.11. In each round, Phil is presented with eight worms, each of which carries a URL that is displayed when Phil moves near it, as shown in Figure 8.12. The player can move Phil around the screen and "eat" the real worms or "reject" the bait. Phil is rewarded with 100 points if he correctly eats a good worm or correctly rejects a bad one. He is slightly penalized for rejecting a good worm (false positive) by losing 10 seconds from the clock for that round. He is severely penalized if he eats a bad worm and is caught by phishers (false negative), losing one of his three lives. Players have to correctly recognize at least six out of eight URLs within two minutes to move on to the next round. As long as they still have lives, they can repeat a round until they are able to recognize at least six URLs correctly. If a player loses all three lives the game is over. At the end of every round, a review screen shows all of the URLs from that round and provides tips for identifying them correctly, as shown in Figure 8.13.

The game is implemented in Flash 8. The content for the game, including URLs and training messages, is loaded from a separate data file at the start of the game. This makes it easy to quickly update the content. In each round of the game, four good worms and four phishing worms are randomly selected from the twenty URLs in the data file for that round. Sound effects are used to provide audio feedback, and background music and underwater background scenes help keep users engaged.

Educational action design methodology is used to design the game. In this method, the learner is given a stipulated time in which they have to perform (and thereby learn) the things that are presented in the game [20]. Table 8.6 summarizes the ways instructional design principles were applied to the design of Anti-Phishing Phil.

[2] http://wombatsecurity.com/



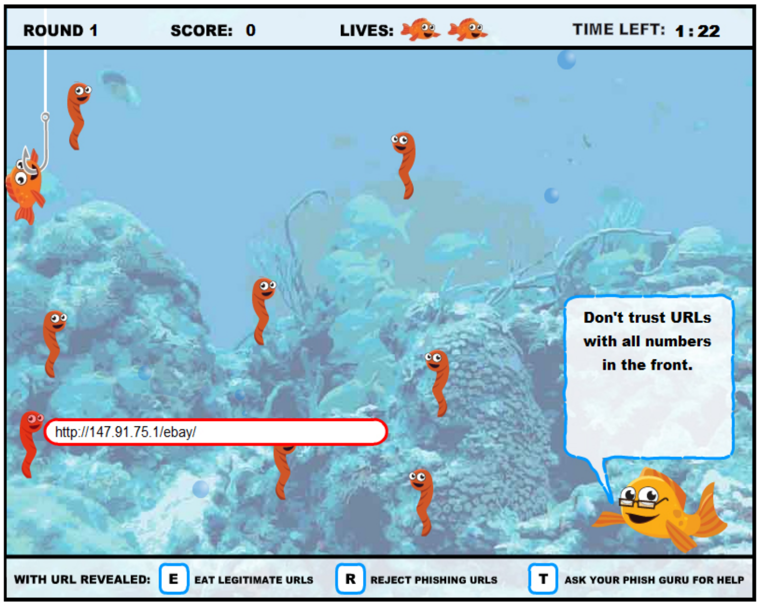
How To Avoid Online Scams

Within the Address, the right-hand side is the most important! It shows the site name.

https://www.wombank.com/wombank.html

I got it, Dad! wombank.com is the site name!

Figure 8.11: Screen shot from Anti-Phishing Phil. The screen shows part of one of the tutorials that occur before the beginning of each round.



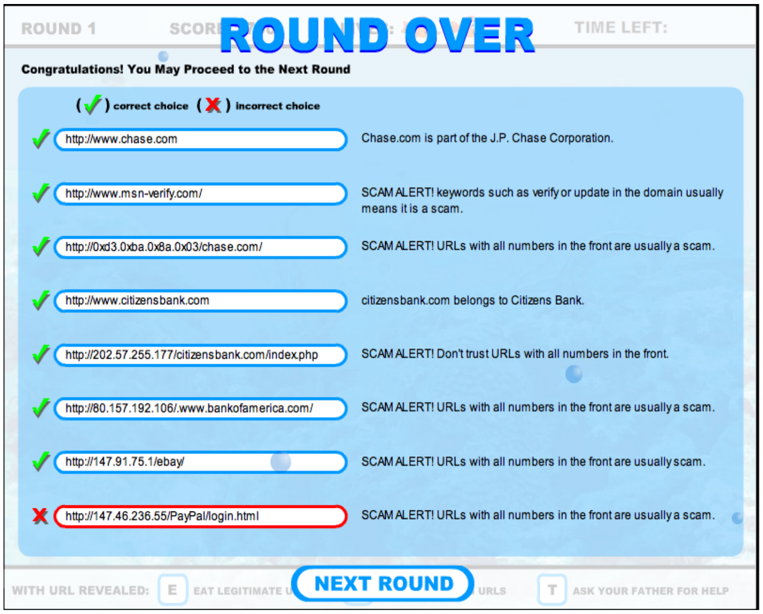
ROUND 1 SCORE: 0 LIVES: TIME LEFT: 1:22

http://147.91.75.1/ebay/

Don't trust URLs with all numbers in the front.

WITH URL REVEALED: E EAT LEGITIMATE URLS R REJECT PHISHING URLS T ASK YOUR PHISH GURU FOR HELP

Figure 8.12: Screen shot from Anti-Phishing Phil. The screen shows a URL being displayed as Phil swims by a worm; the lower right corner features a tip from the PhishGuru fish.



Congratulations! You May Proceed to the Next Round

(✓) correct choice (✗) incorrect choice

✓	http://www.chase.com	Chase.com is part of the J.P. Chase Corporation.
✓	http://www.msn-verify.com/	SCAM ALERT! keywords such as verify or update in the domain usually means it is a scam.
✓	http://0xd3.0xba.0x8a.0x03/chase.com/	SCAM ALERT! URLs with all numbers in the front are usually a scam.
✓	http://www.citizensbank.com	citizensbank.com belongs to Citizens Bank.
✓	http://202.57.255.177/citizensbank.com/index.php	SCAM ALERT! Don't trust URLs with all numbers in the front.
✓	http://80.157.192.106/.www.bankofamerica.com/	SCAM ALERT! URLs with all numbers in the front are usually a scam.
✓	http://147.91.75.1/ebay/	SCAM ALERT! URLs with all numbers in the front are usually scam.
✗	http://147.46.236.55/PayPal/login.html	SCAM ALERT! URLs with all numbers in the front are usually a scam.

Figure 8.13: Screen shot from Anti-Phishing Phil. The screen shows the end of round summary.

8.2.2 Anti-Phishing Phil lab study

Study design

Using the protocol introduced in Section 5.1, we conducted a study to measure how much knowledge participants acquired by playing Anti-Phishing Phil. Participants were asked to examine 10 websites and determine which were phishing websites. After 15 minutes of training, they were asked to examine 10 more websites and determine which were phishing websites. Half of the websites were phishing websites based on popular brands, while the other half were legitimate websites from popular financial institutions, online merchants, and other random sources.

As this research was also focused on educating novice users about phishing attacks, participants with little technical knowledge were recruited. Fliers were posted around our university and local neighborhoods; users were then screened through an online survey. Twenty-eight participants were recruited and assigned randomly to either a “tutorial” condition or “game” condition. In the tutorial condition, participants were asked to spend up to fifteen minutes reading an anti-phishing tutorial based on the Anti-Phishing Phil game. The tutorial included 17 pages of color printouts containing all of the between-round training messages and URL lists used in the game. These lists included explanations of which were legitimate URLs and which were phishing URLs, similar to the game's end-of-round screens. In the game condition, participants played the Anti-Phishing Phil game for fifteen minutes.

The results of the Anti-Phishing Phil lab study were compared with the data from the existing

Table 8.6: Applying the instructional design principles in Phil design.

Principle	Way(s) in which we applied the principle to our design
Learning-by-doing	Users identify real and fake websites while playing a game
Conceptual-procedural	Applied in the between-round tutorials, for example, we provide information about how to search for a brand or domain and how to decide which of the search results are legitimate (procedural knowledge) after mentioning that search engines are a good method to identify phishing websites (conceptual knowledge)
Contiguity	Applied in the between-round tutorials
Personalization	Applied in the messages from the father fish
Story-based agent environment	Applied by having the user control a young fish named Phil (agent), who has to learn anti-phishing skills to survive in the water among sharks and big fishes (story)
Reflection	Applied at the end of each round by displaying a list of websites that appeared in that round and an indication as to whether the user correctly or incorrectly identified each one

training material evaluation presented in Section 5.1. Table 8.7 shows the demographic details of the participants in both studies.

Results

The study measured how much knowledge participants acquired by playing Anti-Phishing Phil. It did so by examining false positives, false negatives, and the total percentage of correct websites identified before and after playing the game. A false positive occurs when a legitimate website is mistakenly judged to be a phishing website. A false negative occurs when a phishing website is incorrectly judged to be a legitimate website. As shown in Figure 8.14, the game condition performed best overall. It performed roughly as well as the existing training material condition in terms of false negatives, and better on false positives. The tutorial condition also performed better than the existing training material in terms of false positives, but this was not statistically significant.

Post-test false negative rates in all three groups decreased significantly from the pre-test values. For the existing training materials condition, the false negative rate fell from 0.38 to 0.12 (paired t-test, p-value = 0.01). For the tutorial condition, it changed from 0.43 to 0.19 (paired t-test, p-value < 0.03). For the game condition, it changed from 0.34 to 0.17 (paired t-test, p-value < 0.02). There was no statistical difference between the groups in either the pre-test (oneway ANOVA, p-value

Table 8.7: Participants for the Anti-Phishing Phil study.

Characteristics	**Conditions**			
	Existing training material	**Tutorial**	**Game**	**Control**
Sample size	14	14	14	14
Gender				
Male	29%	36%	50%	33%
Female	71%	64%	50%	67%
Age				
18 - 34	93%	100%	100%	93%
> 34	7%	0%	0%	7%
Education				
High School	14%	7%	7%	9%
College Undergrad	51%	79%	51%	48%
College graduate	14%	7%	21%	22%
Post. Graduate school	21%	7%	21%	22%
Years on the Internet				
3- 5 years	23%	23%	15%	15%
6-10 years	69%	70%	78%	70%
> 11 years	8%	7%	7%	15%

= 0.60), or post-test (oneway ANOVA, p-value = 0.45). Post-test false positive rates decreased significantly in the game condition (p-value < 0.03). A one-way ANOVA revealed that false positive rates differed significantly in the post-test (paired t-test, p-value < 0.02). The Tukey post-hoc test revealed that the game condition had significantly lower false positives than the existing training materials. No other specific post-hoc contrasts were significant.

The results demonstrate that users showed significant improvements in their ability to identify phishing links correctly after 15 minutes of training with Anti-Phishing Phil, the tutorial, or existing online training materials. However, participants in the game condition were better able to distinguish between phishing and legitimate links than those in the other conditions, and were thus less likely to incorrectly identify legitimate links as phishing links.

8.2.3 Anti-Phishing Phil field study

In this section, we discuss results from data collected in a real-world deployment of Anti-Phishing Phil. Results provide more evidence that Anti-Phishing Phil is effective for knowledge acquisition and knowledge retention [111].

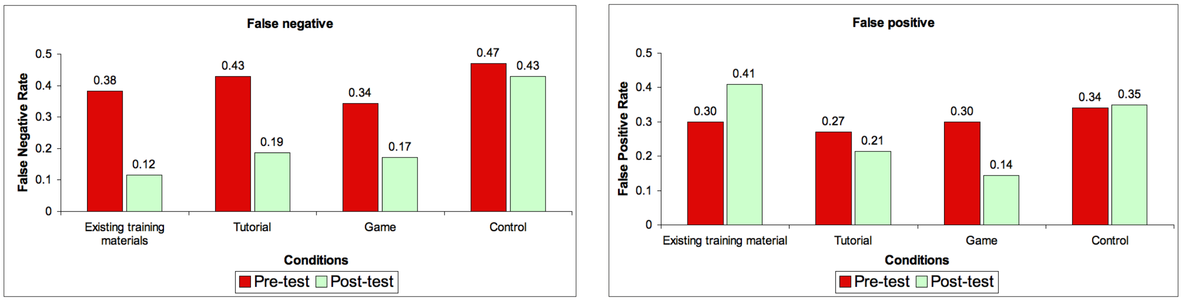


Figure 8.14: False negatives and false positives on pre-test and post-test. The differences in false negatives between groups are not statistically significant. The game condition has significantly lower false positives than the existing training materials.

Study design

Participants were recruited for an online study through online mailing list postings offering participants a chance to win a raffle for a $100 Amazon gift certificate. A between-subjects design was used to test two conditions. In the control condition, participants saw 12 websites and were asked to identify whether each website was a phishing site or not. After doing this, the participants were taken to the game. In the game condition, participants were shown six websites before playing the game (pre-test) and another six websites after they finished playing the game (immediate post-test). To measure retention, we emailed participants seven days later and asked them to take a similar test (delayed post-test). In total, each participant in the game condition was tested on 18 websites divided into three groups with each group containing three phishing websites and three legitimate websites. The order of websites within each group and the order in which the groups were shown to each participant was randomized.

Participants

Over the course of two weeks (Sep 25, 2007 to Oct 10, 2007), 4,517 people participated in the study. In the game condition, 2,021 users completed both the pre-test and immediate post-test, 674 of whom came back one week later for the delayed post-test. In this analysis, we focus on people who completed the pre-test, immediate post-test, and delayed post-test. We had 2,496 participants in the control condition. Among the total participants, 78% were male, and 15.6% female, with 6.4% declining to give their gender; 4.8% were 13 - 17 years old, 43.7% 18 - 34 years old, 44.3% 35 - 64 years old, and 0.5% more than 65 years old, with 6.8% declining to provide their age.

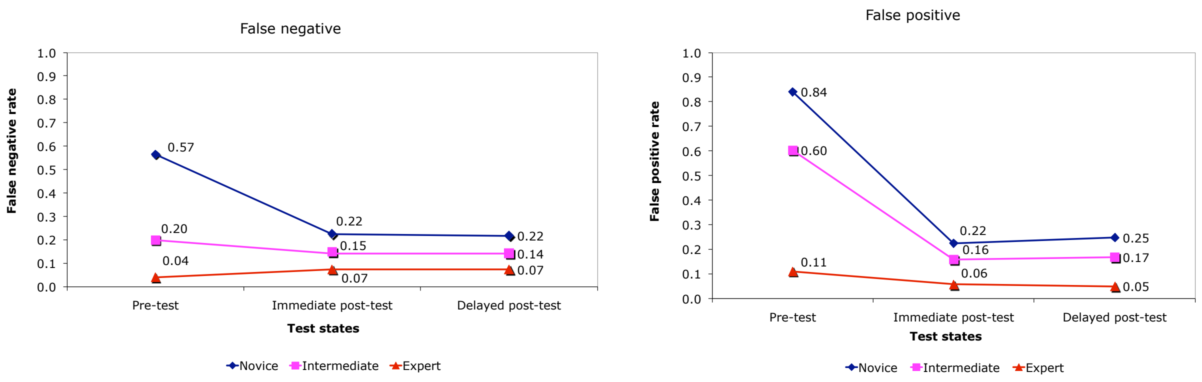


Figure 8.15: False negative and false positive rate for Anti-Phishing Phil in the real-world. Novice users showed the greatest improvement in false negative and false positive rates.

Results

The results demonstrated that users are able to more accurately and quickly distinguish phishing websites from legitimate websites after playing the game, and that users retain knowledge learned from the game for at least one week.

The game condition participants were classified into three categories based on their pre-test scores: novice (0 - 2 correct), intermediate (3 - 4 correct) and expert (5 - 6 correct). As illustrated in Figure 8.15, novice users showed the greatest improvement, with the false positive rate decreasing from 42% to 11.2% (paired t-test, p-value < 0.0001) and the false negative rate decreasing from 28.3% to 11.2% (paired t-test, p-value < 0.0001). The intermediate group also showed statistically significant improvement, though it was not as large as the novice group. Finally, we did not observe any statistically significant improvement in the expert group. Delayed post-test scores did not decrease from immediate post-test scores, demonstrating that participants retained their knowledge after one week.

Participants were able to determine website legitimacy more quickly after playing the game. The mean time users in the game group took to determine a website's legitimacy before the game was 21.2 seconds. After the game, it decreased to 11.2 seconds (paired t-test, p-value < 0.0001). The mean scores for the control group did not change in a statistically significant way (pre - 18.5 seconds, post - 18.6 seconds).

Those who did not come back for the delayed post-test performed slightly worse than those who did come back. The immediate post-test score was 83.8% for those who did not come back and 89.1% for those who did come back one week later (two sample t-test, $p < 0.001$). One possible explanation is that those who were more confident in their performance were more likely to come back. To validate this hypothesis, we conducted a Chi-square test of the percentage of novice, intermediate and expert users who completed the immediate post-test, or delayed post-test. We found that there were more experts and fewer intermediate and novices in the delayed post-test group ($p < 0.001$).

Before playing the game, the mean accuracy scores for males were significantly higher than those for females (males = 75.5%, females = 64.4%, two sample t-test, $t = 8.48$, $p < 0.0001$). However, the two groups improved similarly after playing the game (two proportion test, 14.2% versus 12.4%, $p = 0.192$). There was also a significant difference in pre-test performance between different age groups (one way ANOVA $F = 7.29$, $p < 0.01$). A Tukey simultaneous 95% confidence interval test revealed that participants whose age was less than 18 performed worse than those between 18 and 64 years old. There was no statistical difference in performance between the age groups 18-35 and 36-64. We observed similar trends in immediate post-test performance (one way ANOVA, $F = 23.05$, $p < 0.01$). These results suggested that teenagers may be particularly susceptible to phishing attacks. The mean scores for the age group 13-17 years was 3.9 while the mean score was 4.6 for both the 18 - 34 and 35 - 64 age groups.

The data from the game was used to determine which types of URLs are most difficult for people to identify correctly. Especially challenging URLs included those longer than the address bar and deceptive URLs that look similar to legitimate URLs but with added text (e.g. http://www.msn-verify.com/). The more challenging the URL, the more likely game players were to use the game's help feature ($r = -0.645$, $p < 0.001$). From the game data, it was found that users were most confused by long URLs. This confusion makes them susceptible to sub-domain attacks such as (https://citibusinessonline.da-us.citibank.com/cbusol/signon.do). Users are also confused by very similar URLs. For example, www.citicards.net (as opposed to www.citicards.com), www.eztrade.com (as opposed to www.etrade.com). Further investigation should explore ways to alleviate this confusion among users.

8.2.4 Discussion

Security education plays an important role in increasing users' alertness to security threats. Alert users are cautious, and therefore less likely to make mistakes that will leave them vulnerable to attack (false negatives). However, cautious users tend to misjudge non-threats as threats (false positives) unless they have learned how to distinguish between the two. Good security user education should not only increase users' alertness, but also teach them how to distinguish threats from non-threats. In this section, signal detection theory (SDT) [119,169] is used to quantify the ability to discern between signal (phishing websites) and non-signal or noise (legitimate websites).

Two measures–*sensitivity* (d') and *criterion* (C)–are used in the user studies. Sensitivity is defined as the ability to distinguish phishing websites from legitimate websites; it is measured by the distance between the mean of signal and non-signal distributions. The larger the value of d', the

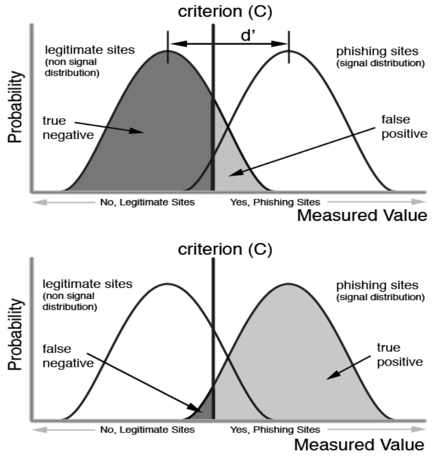


Figure 8.16: Applying signal detection theory (SDT) to anti-phishing education. Legitimate websites are treated as "non-signal," and phishing websites as "signal." Sensitivity (d') measures users' ability to distinguish signal from non-signal. Criterion (C) measures users' decision tendency (C < 0 indicates cautious users , C = 0 indicates neutral users, C > 0 indicates liberal users). As a result of training users may a) become more cautious, increasing C; b) become more sensitive, increasing d'; or c) a combination of both.

better the user is at separating signal from noise. *Criterion* is defined as the tendency of users to exercise caution when making a decision. More cautious users are likely to have few false negatives and many false positives, while less cautious users are likely to have many false negatives and few false positives. Figure 8.16 shows example distributions of user decisions about legitimate and phishing websites. The criterion line divides the graph into four sections representing true positives, true negatives, false positives, and false negatives. Training may cause users to become more cautions, increasing C and moving the criterion line to the right. Alternatively, training may cause users to become more sensitive, separating the two means. In some cases, training may result in both increased caution and increased sensitivity, or in decreased caution but increased sensitivity.

C and d' were calculated for the evaluation of existing online training materials, the Anti-Phishing Phil laboratory study, and the Anti-Phishing Phil field study, as summarized in Table 8.8. It was found that, after reading existing training materials, users became significantly more cautious without becoming significantly more sensitive. Thus, these materials serve to increase alertness, but do not teach users how to distinguish legitimate websites from fraudulent ones. After playing Anti-Phishing Phil, users became significantly more sensitive and liberal, indicating that performance improvements from playing the game were due to learning. (Note: we observed the Criterion change in the field study and not in the laboratory study.)

Results from the Anti-Phishing Phil studies demonstrate that participants who played the game were better able to identify phishing websites than participants who completed two other types of training. In the evaluation of both approaches, it was found that people could retain what they learned for at least one week without significant degradation in performance.

Table 8.8: Signal Detection Theory analysis. Anti-Phishing Phil increased users' sensitivity, while existing training materials made users more cautious. * indicates statistically significant differences (p <0.05).

	Sensitivity (d')			**Criterion (C)**		
	Pre-test	**post-test**	**Delay**	**Pre-test**	**post-test**	**Delay**
Existing training materials	0.81	1.43	–	0.03	-0.51*	–
Anti-Phishing Phil laboratory study	0.93	2.02*	–	0.06	0.06	–
Anti-Phishing Phil field study	1.49	2.46*	2.47	-0.35	0.02*	0.0

Chapter 9

Conclusions, Recommendations and Future Work

In this chapter, we present conclusions from this thesis work, insights from our research, some recommendations for security education, and future plans we have in this line of research.

9.1 Conclusions

In this thesis, we have systematically studied the problem of educating users about phishing (a semantic attack). Through well-designed studies, we have shown that users can be trained effectively if training materials are presented when users "fall" for phishing attacks. Through this thesis work, we have created effective training interventions and developed a novel approach (delivery mechanism) for their presentation. With PhishGuru, we address the three challenges of security user education by: (1) motivating users to read the training interventions; (2) making training part of the primary task itself (through emails); and (3) ensuring that PhishGuru training does not increase users' tendency to misjudge non-threats as threats. In this thesis work, we have also developed and evaluated interventions grounded in learning science principles. We evaluated both the delivery mechanism and content through laboratory and real world studies. PhishGuru effectively trains people; furthermore, people trained with PhishGuru retain knowledge even after four weeks. We believe the success of PhishGuru is due to the fact that its methodology and content are grounded in theory.

Through laboratory and real-world studies, we have shown that:

> **Computer users trained using an embedded training system grounded in learning science theory are able to make more accurate online trust decisions**

than those who read traditional security training materials distributed via email or posted on web sites.

The results of this thesis are not only applicable to education specifically centered on phishing, but also security education in general. The design principles established in this thesis will help researchers develop systems that can train users in other risky online situations.

9.2 Recommendations for security education

Researchers tend to agree that no system will ever be completely accurate at detecting phishing attacks, especially when detection requires contextual information. By training users to make better decisions, we offer a complementary approach that can be put into immediate practice. Based on the lessons learned from the laboratory and real-world studies, the following design principles should inform the design of any security user-education system:

- **Integrate security education into users' primary tasks.** For most users, security is a secondary task (e.g. one does not go to an online banking website to check the SSL implementation of the website, but rather to perform a banking transaction). Also, since users are not motivated to read about security in general, they do not take the time to educate themselves about security. Therefore, making education part of a primary task is essential to motivating people to read training materials. People reading training materials as part of the primary task may remember the instructions better than people who read the training materials in an isolated fashion.

- **Interventions should apply instructional design principles.** Educational researchers have developed and evaluated instructional design principles; these principles should be applied to the design of interventions and training materials. Some principles are very easy to apply and can quite effectively help people remember instructions. Those designing instructions in the future should consider the following principles in particular:

    - **Present the instructions in a comic strip format.** One reason PhishGuru has been effective is that the instructions are presented in a comic strip format. Participants in all of our studies mentioned that instructions presented in a comic strip format are very effective and likely to be read by people of all ages. Therefore, we think the comic strip format should be used to develop security training materials. However, it may be worth testing other formats, such as video.
    - **Make the training materials fun and interactive.** Another reason both PhishGuru and Phil have been effective is that the training materials are presented in a fun and

  interactive manner. Both PhishGuru and Phil use stories populated by characters. Phil is more interactive in the sense that users have to do certain things in order to learn; in PhishGuru, the interaction between the learner and the intervention is through text presented in the form of dialogue between characters. Based on our research and the results of our studies, future developers of security training material should make training more fun and interactive for the users.

  - **Present the instructions in a story format.** PhishGuru and Phil have been effective because both systems make use of an underlying story. In the PhishGuru intervention, PhishGuru tries to stop the victim from falling for phishing by presenting tips on how to protect oneself and information about how phishers scam victims. In the Anti-Phishing Phil game, Phil learns how to identify phishing URLs through PhishGuru. Based on our research, training materials using a narrative format effectively help users understand and retain important information.

- **Format the instructions as a list of actionable items.** People don't want to have to read long text blocks to find important information, so format instructions as a list of actionable items (procedural knowledge). Users tend to learn these actionable items and remember them for at least a few weeks.

- **Make the training repetitive.** According to results from this thesis, people who are trained twice do better than people trained only once. This suggests that security training education should be repeated every once in a while.

- **Keep the training messages short and simple.** In our studies, traditional security notices fared poorly because they contained too much text and technical jargon. In both PhishGuru and Phil, instructions are limited to one page and presented succinctly. Based on our research, security training materials should be kept short and simple.

9.3 Future work

- **Apply embedded training in other scenarios.** Throughout this thesis work, we have applied the embedded training concept to materials exclusively about Phishing. However, this methodology can be extended to many other interesting scenarios, embedding training materials into things like an instant message link or email attachment. Future research should work to determine the effectiveness of embedded training in other scenarios.

- **Test other mediums of training.** Throughout this thesis, we have used still images in the PhishGuru intervention. We believe that it would be worthwhile to investigate if other media–such as a short narrative video—might be more effective than still images.

- **Study longer retention and the effect of more training.** In this thesis, the maximum retention time studied was 28 days. It would be interesting to study longer periods of retention, like 6 months. We have also only studied the effect of one and two training messages. We found that, even after people saw two training messages, many still fell for phishing attacks. It would be interesting to study the effect of more training messages on peoples' tendency to fall for attacks. In addition, it would be worthwhile to study why people fall for phishing attacks even after they see the training interventions twice. It would also be worthwhile to investigate how often people should be trained for training to be most effective.

- **Build a system that can automate the entire process from email creation to data analysis.** As part of this thesis, we have built a prototype of PhishGuru that is semi-automatic in nature. To aid the process of setting up an implementation of PhishGuru and beginning data collection, it would be useful to develop a system that is completely automatic.[1]

- **Leverage PhishGuru to convince people for more training.** In this thesis, we used the teachable moment – the moment when users click on a link and fall for fake phishing emails – to present training materials to users. We believe that this moment can also be used to convince users to sign up for more extensive training on phishing and other security-related concepts.

- **Cost-benefit analysis.** The costs of PhishGuru are three-fold: one development and implementation of the PhishGuru infrastructure along with the interventions; second, the time investment of end users; and third, the analysis of collected data. The main benefit (from all of the studies we have done) is around 50% improvement in the behavior (identifying phishing emails correctly) of the users. It will be worthwhile to develop an economic cost-effectiveness model for PhishGuru.

[1] I am currently helping Wombat Security Technologies build the entire system.

Bibliography

[1] Account Guard. Account Guard, 2006. Retrieved Nov 3, 2006, http://pages.ebay.com/ebay_toolbar/.

[2] Adams, A., and Sasse, M. A. Users are not the enemy. *Communication ACM 42*, 12 (1999), 40–46. DOI=http://doi.acm.org/10.1145/322796.322806.

[3] Akerlof, G. A. The market for lemons: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics 84*, 3 (1970), 488–500.

[4] Aldrich, C. *Learning by Doing: A Comprehensive Guide to Simulations, Computer Games, and Pedagogy in e-Learning and Other Educational Experiences*. Jossey-Bass, 2005.

[5] Aleven, V. An intelligent learning environment for case-based argumentation. *Technology, Instruction, Cognition, and Learning 4*, 2 (2006), 191 – 241.

[6] Aleven, V., and Koedinger, K. R. An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science 26*, 2 (2002), 147 – 179.

[7] Allen, M. Social engineering: A means to violate a computer system. Tech. rep., SANS Institute, 2006.

[8] American National Standards Institute. Environmental and facility safety signs.

[9] Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., and Roinestad, H. Phishing IQ tests measure fear, not ability. *Usable Security (USEC'07)* (2007). http://usablesecurity.org/papers/anandpara.pdf.

[10] Anderson, J. R. *Rules of the Mind*. Lawrence Erlbaum Associates, Inc., 1993.

[11] Anderson, J. R., Boyle, C. F., Corbett, A. T., and Lewis, M. W. Cognitive modelling and intelligent tutoring. *Artificial Intelligence 42* (1990), 7–49.

[12] Anderson, J. R., Corbett, A. T., Koedinger, K. R., and Pelletier, R. Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences 4*, 2 (1995), 167–207.

[13] Anderson, J. R., and Simon, H. A. Situated learning and education. *Educational Researcher 25* (1996), 5–11.

[14] Ang, L., Dubelaar, C., and Lee, B. C. To trust or not to trust? a model of internet trust from the customer's point of view. In *Proceedings, 14th Bled Electronic Commerce Conference* (June 2001), pp. 25–26.

[15] Anti-Phishing Working Group. Anti-Phishing Working Group, 2007. http://www.antiphishing.org/.

[16] Araujo, I., and Araujo, I. Developing trust in internet commerce. In *Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research* (2003), IBM Press, pp. 1–15. Retrieved Sept 13, 2005, http://portal.acm.org/citation.cfm?id=961324.

[17] Artz, D., and Gil, Y. A survey of trust in computer science and the semantic web. *Web Semant. 5*, 2 (2007), 58–71.

[18] Asgharpour, F., Liu, D., and Camp, L. J. Mental models of computer security risks. *Workshop on the Economics of Information Security* (2007).

[19] Bahrick, H. P. Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology 108*, 3 (September 1979), 296–308.

[20] Baker, R., Habgood, J., Ainsworth, S. E., and Corbett, A. Modeling the acquistion of fluent skill in educational action games. *Proceedings of User Modeling 2007* (2007), 12–26.

[21] Bank, D. 'Spear phishing' tests educate people about online scams. News article, Augurst 2005. http://online.wsj.com/public/article/SB112424042313615131-z_8jLB2WkfcVtgdAWf6LRh733sg_20060817.html?mod=blogs.

[22] Bank of America. Reactive phishing education: An industry solution to phishing. Tech. rep., Bank of America, 2007.

[23] Barnett, S. M., and Ceci, S. J. When and where do we apply what we learn? a taxonomy for far transfer. In *Psychological Bulletin* (2002), vol. 128, pp. 612–637.

[24] Bellman, S., Johnson, E. J., Kobrin, S. J., and Lohse, G. L. International Differences in Information privacy concerns: A global survey of consumers. *The Information Society 20* (2004), 313 – 324.

[25] Bhattacherjee, A. Individual Trust in Online Firms: Scale Development and Initial Test. *Journal of Management Information Systems 19*, 1 (2002), 211–242.

[26] Bishop, M. Education in information security. *IEEE Concurrency 8*, 4 (2000), 4–8.

[27] Bligh, D. A. *What's The Use of Lectures?* Jossey-Bass, 2000.

[28] Bransford, J. D., and Schwartz, D. L. Rethinking transfer: A simple proposal with multiple implications. In *Review of Research in Education*, A. Iran-Nejad and P. D. Pearson., Eds., vol. 24. American Educational Research Association (AERA) Washington, DC, 2001, pp. 61 – 100.

[29] Brewer, M. *Handbook of Research Methods in Social and Personality Psychology.* Cambridge University Press, 2000.

[30] Burmester, G. M., Stottler, D., and Hart, J. L. Embedded training intelligent tutoring systems (ITS) for the future combat systems (FCS) command and control (C2) vehicle. Tech. rep., Defense Technical Information Center, 2005. http://www.stottlerhenke.com/papers/IITSEC-02-ITSFCS.pdf.

[31] Burt, R. S. Social contagion and innovation: Cohesion versus structural equivalence. *The American Journal of Sociology 92*, 6 (1987), 1287–1335.

[32] Butcher, K. R., and Aleven, V. Integrating visual and verbal knoweldge during classroom learning with computer tutors. To appear in Cognitive Science.

[33] Cabral, L. M. B. The Economics of Trust and Reputation: A Primer. Tech. rep., New York University and CEPR, May 2002. Retrieved Feb 20, 2006, http://pages.stern.nyu.edu/ lcabral/reputation/Reputation_June05.pdf.

[34] Camp, J. L. Mental models of privacy and security. *IEEE Technology and Society* (2007).

[35] Camp, J. L. Mental models of privacy and security, 2007. Retrieved Oct 10, 2007, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=922735.

[36] Cave, J. The economics of cyber trust between cyber partners. Tech. rep., Cyber Trust & Crime Prevention Project, April 2004. Retrieved Dec 11, 2007, http://www.berr.gov.uk/files/file15291.pdf.

[37] Cavoukian, A., and Hamilton, T. *The Privacy Payoff, How Successful Business Build Consumer Trust.* McGraw Hill Tyerson Limited, 2002.

[38] Chase, W. G., and Simon, H. A. Perception in chess. *Cognitive Psychology 4* (1973), 55 – 81.

[39] Chellappa, R. K. Consumers' Trust in Electronic Commerce Transactions: The Role of Perceived Privacy and Perceived Security. Tech. rep., 2005. Under review. Retrieved Sept 13, 2005, http://asura.usc.edu/ ram/rcf-papers/sec-priv.pdf.

[40] Chellappa, R. K., and Sin, R. Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. *Information Technology and Management. Vol. 6, No. 2-3* (2005.). Retrieved Sept 13, 2005, http://asura.usc.edu/ ram/rcf-papers/per-priv-itm.pdf.

[41] Chi, M. T. H., Feltovich, P., and Glaser, R. Categorization and representation of physics problems by experts and novices. *Cognitive Science 5* (1981), 121–152.

[42] Cialdini, R. B. The science of persuasion. *Scientific American* (February 2001).

[43] Clark, R. C. *Developing Technical Training: A Structured Approach for the Development of Classroom and Computer-Based Instructional Materials.* Addison Wesley Publishing Company, Beverly, MA, USA, June 1989.

[44] Clark, R. C., and Mayer, R. E. *E-Learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning.* John Wiley & Sons, Inc., USA, 2002.

[45] Coates, R. Dumb users spread viruses. Retrieved Jan 24, 2007, http://www.silicon.com/software/security/0,39024655,39118228,00.htm.

[46] Committee on Developments in the Science of Learning and National Research Council. *How People Learn: Bridging Research and Practice.* National Academies Press, 2000.

[47] Corbett, A. T., and Anderson, J. R. Locus of feedback control in computer-based tutoring: impact on learning rate, achievement and attitudes. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2001), ACM Press, pp. 245–252.

[48] Cordova, D. I., and Lepper, M. R. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology 88*, 4 (December 1996), 715–730.

[49] Corritore, C. L., Kracher, B., and Wiedenbeck, S. On-line trust: concepts, evolving themes, a model. *Academic Press, Inc. 58*, 6 (2003), 737–758.

[50] Cranor, L. F. A framework for reasoning about the human in the loop. *Usability, Psychology, and Security* (2008).

[51] Cranor, L. F., and Garfinkel, S. *Security and Usability: Designing Secure Systems that People Can Use.* O'Reilly, Sebastopol, CA, USA, Aug, 2005.

[52] Dhamija, R., and Tygar, J. The Battle Against Phishing: Dynamic Security Skins. In *SOUPS '05: Proceedings of the 2005 symposium on Usable privacy and security* (New York, NY, USA, 2005), ACM Press, New York, NY, pp. 77–88. Retrieved Feb 10, 2006, DOI= http://doi.acm.org/10.1145/1073001.1073009.

[53] Dhamija, R., Tygar, J. D., and Hearst, M. Why Phishing Works. *Proceedings of the Conference on Human Factors in Computing Systems (CHI)* (2006).

[54] Downs, J., Holbrook, M., and Cranor, L. Decision Strategies and Susceptibility to Phishing. In *SOUPS '06: Proceedings of the second symposium on Usable privacy and security* (New York, NY, USA, 2006), ACM Press, pp. 79–90.

[55] Downs, J., Holbrook, M., and Cranor, L. Behavioral response to phishing risk. In *eCrime '07: Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit* (New York, NY, USA, October 2007), ACM, pp. 37–44.

[56] eBay. Spoof email tutorial, 2006. Retrieved December 30, 2006. http://pages.ebay.com/education/spooftutorial.

[57] Eberts, R. E. *Handbook of Human-computer Interaction.* Elsevier Science, 1997, ch. Computer-based Instruction, pp. 825–847.

[58] Egelman, S., Cranor, L. F., and Hong, J. You've been warned: An empirical study of the effectiveness of web browser phishing warnings. CHI '08: Proceedings of the SIGCHI conference on Human factors in computing systems.

[59] Ellis, E., Worthington, L., and Larkin, M. Research synthesis on effective teaching principles and the design of quality tools for educators. Tech. rep., National center to improve the tools of educators, 1994.

[60] Emigh, A. Online identity theft: Phishing technology, chokepoints and countermeasures. Tech. rep., Radix Labs, October 2005. http://www.antiphishing.org/Phishing-dhs-report.pdf.

[61] Evers, J. User education is pointless, October 2006. http://news.com.com/2100-7350_3-6125213.html.

[62] Federal Trade Commission. An e-card for you game, 2006. Retrieved December 30, 2006. http://www.ftc.gov/bcp/conline/ecards/phishing/index.html.

[63] Federal Trade Commission. How not to get hooked by a phishing scam. Consumer alert news, 2006. Retrieved Nov 7, 2006, http://www.ftc.gov/bcp/edu/pubs/consumer/alerts/alt127.htm.

[64] Ferguson, A. J. Fostering E-Mail Security Awareness: The West Point Carronade. *EDUCASE Quarterly*, 1 (2005). Retrieved March 22, 2006, http://www.educause.edu/ir/library/pdf/eqm0517.pdf.

[65] Fette, I., Sadeh, N., and Tomasic, A. Learning to detect phishing emails. *16th International conference on World Wide Web* (June 2006). Retrieved Sept 2, 2006, http://reports-archive.adm.cs.cmu.edu/anon/isri2006/CMU-ISRI-06-112.pdf.

[66] Financial Services Technology Consortium. Understanding and countering the phishing threat: A financial services industry perspective. Tech. rep., Financial Services Technology Consortium, 2005.

[67] Florencio, D., and Herley, C. Stopping a phishing attack, even when the victims ignore warnings. Tech. rep., Microsoft, 2005.

[68] Fogg, B. *Persuasive Technology: Using Computers to Change What We Think and Do.* Morgan Kaufmann, December 2002.

[69] Fong, G. T., and Nisbett, R. E. Immediate and delayed transfer of training effects in statistical reasoning. In *American Psychological Association Inc.*, vol. 120. Journal of Experimental Psychology, 1991, pp. 34–45.

[70] Frantz, J. P., and Rhoades, T. P. A task-analytic approach to the temporal and spatial placement of product warnings. *Human Factors: The Journal of the Human Factors and Ergonomics Society 35*, 4 (December 1993), 719 – 730.

[71] Frederick, S. Cognitive reflection and decision making. *Journal of Economic Perspectives 19*, 4 (2005), 25–42.

[72] Gagne, R. M., Foster, H., and Crowley, M. E. The measurement of transfer of training. *Psychological Bulletin 45*, 2 (1948), 97–130.

[73] Garretson, C. Whaling: Latest e-mail scam targets executives. News article, November 2007. Retrieved Dec 17, 2007, http://www.networkworld.com/news/2007/111407-whaling.html.

[74] Gartner. Gartner Survey Shows Frequent Data Security Lapses and Increased Cyber Attacks Damage Consumer Trust in Online Commerce. Tech. rep., June 2005. Retrieved Jan 9, 2007, http://www.gartner.com/press_releases/asset_129754_11.html.

[75] Gefen, D. Reflections on the dimensions of trust and trustworthiness among online consumers. *ACM Press 33*, 3 (2002), 38–53. DOI=http://doi.acm.org/10.1145/569905.569910.

[76] Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., and Soutter, C. L. Measuring trust. *The Quarterly Journal of Economics 115*, 3 (2000), 811–846.

[77] Gobet, F., and Simon, H. A. Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology 31*, 1 (August 1996), 1–40.

[78] Goldberg, L. R. Man versus model of man: a rationale, plus some evidence for a method of improving clinical judgment. *Psychological bulletin 73* (1970), 422 – 432.

[79] Gordon, L. A., Loeb, M. P., Lucyshyn, W., and Richardson, R. CSI/FBI Computer Crime and Security Survey. Report, Computer Security Institute, 2006.

[80] Gorling, S. The myth of user education. In *Proceedings of the 16th Virus Bulletin International Conference* (2006).

[81] Grabner, S., and Kaluscha, E. A. Empirical research in on-line trust: a review and critical assessment. *Academic Press, Inc 58*, 6 (2003), 783–812.

[82] Granger, S. Social engineering fundamentals, part I: Hacker tactics. News article, December 2001. Retrieved May 9, 2008, http://www.securityfocus.com/infocus/1527.

[83] Granger, S. Social engineering fundamentals, part II: Combat strategies. News article, January 2002. Retrieved May 9, 2008, http://www.securityfocus.com/infocus/1533.

[84] Guerra, G. A., Zizzo, D. J., Dutton, W. H., and Peltu, M. Economics of Trust in the Information Economy: Issues of Identity, Privacy and Security. Tech. rep., Oxford Internet Institute, April 2003. Retrieved Feb 20, 2006, http://www.oii.ox.ac.uk/resources/publications/RR1.pdf.

[85] Hellier, E., Wright, D. B., Edworthy, J., and Newstead, S. On the stability of the arousal strength of warning signal words. 577 – 592.

[86] Hight, S. D. The importance of a security, education, training and awareness program, November 2005. http://www.infosecwriters.com/text_resources/pdf/SETA_SHight.pdf.

[87] Hiner, J. Change your company's culture to combat social engineering attacks. News article, May 2002. Retrieved Nov 3, 2006, http://articles.techrepublic.com.com/5100-10878_11-1047991.html.

[88] ISO/IEC. ISO/IEC 27001:2005 - information technology – security techniques – information security management systems – requirements. Tech. rep., International Organization for

Standardization (ISO) and the International Electrotechnical Commission (IEC), October 2005.

[89] Jackson, C., Simon, D., Tan, D., and Barth, A. An evaluation of extended validation and picture-in-picture phishing attacks. In *Usable Security (USEC'07)* (2007). http://usablesecurity.org/papers/jackson.pdf.

[90] Jagatic, T., Johnson, N., Jakobsson, M., and Menczer, F. Social phishing. *Communications of the ACM 50*, 10 (October 2007), 94–100. Retrieved March 7, 2006, http://www.indiana.edu/ phishing/social-network-experiment/phishing-preprint.pdf.

[91] Jakobsson, M. The human factor in phishing. In *Privacy & Security of Consumer Information* (2007). http://www.informatics.indiana.edu/markus/papers/aci.pdf.

[92] Jakobsson, M., Juels, A., and Ratkiewicz, J. Remote harm-diagnostics. Retrieved, Jan 14, 2007, http://www.ravenwhite.com/files/rhd.pdf.

[93] Jakobsson, M., and Myers, S., Eds. *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft.* Wiley-Interscience, 2006.

[94] Jakobsson, M., and Ramzan, Z. *Crimeware: Understanding New Attacks and Defenses.* Symantec Press, 2008.

[95] Jakobsson, M., and Ratkiewicz, J. Designing Ethical Phishing Experiments: A study of (ROT13) rOnl query features. In Proceedings of the 15th annual World Wide Web Conference (WWW2006).

[96] James, L. *Phishing Exposed.* Syngress Publishing, Canada, November 2005.

[97] Jarvenpaa, S. L., Tractinsky, N., Saarinen, L., and Vitale, M. Consumer trust in an internet store: A cross-cultural validation. *Journal of Computer-Mediated Communciation 5*, 2 (December 1999). Retrieved Sept 19, 2006, http://jcmc.indiana.edu/vol5/issue2/jarvenpaa.html.

[98] Jarvenpaa, S. L., Tractinsky, N., and Vitale, M. Consumer trust in an internet store. *Inf. Tech. and Management 1*, 1-2 (2000), 45–71.

[99] Johnson, B. R., and Koedinger, K. R. Comparing instructional strategies for integrating conceptual and procedural knowledge. In *Proceedings of the Annual Meeting [of the] North American Chapter of the International Group for the Psychology of Mathematics Education* (October 2002), vol. 1–4, pp. 969–978.

[100] Karlof, C., Tygar, J., and Wagner, D. A user study design for comparing the security of registration protocols. *Usability, Psychology, and Security* (2008).

[101] Keinan, G. Decision making under stress: scanning of alternatives under controllable and uncontrollable threats. *Journal of personality and social psychology 52*, 3 (1987), 639–644.

[102] Kirda, E., and Kruegel, C. Protecting users against phishing attacks. *The Computer Journal 49*, 5 (January 2006), 554–561.

[103] Kirkley, J. R., and et al. Problem-based embedded training: An instructional methodology for embedded training using mixed and virtual reality technologies. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)* (2003). http://www.iforces.org/downloads/problem-based.pdf.

[104] Klein, G. *Sources of power : How people make decisions?* The MIT Press Cambridge, Massachusetts The MIT Press, Cambridge, Massachusetts, London, England, February 1999.

[105] Klein, G., Steve, W., Laura, M., and Caroline, Z. Characteristics of skilled option generation in chess. *Organizational Behavior and Human Decision Processes 62*, 1 (April 1995), 63–69.

[106] Koedinger, K. R. Toward evidence for instruction design principles: Examples from cognitive tutor math 6. *Proocedings of the Annual Meeting, Norh American Chapter of the International Group for the Psychology of Mathematics Education 1 – 4* (2002).

[107] Kumaraguru, P., Acquisti, A., and Cranor, L. Trust modeling for online transactions: A phishing scenario. In *Privacy Security Trust* (2006).

[108] Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., and Pham., T. School of phish: A real-world evaluation of anti-phishing training. Under review.

[109] Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., and Nunge, E. Protecting people from phishing: the design and evaluation of an embedded training email system. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2007), ACM Press, pp. 905–914.

[110] Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L. F., and Hong, J. Getting users to pay attention to anti-phishing education: Evaluation of retention and transfer. *e-Crime Researchers Summit, Anti-Phishing Working Group* (2007).

[111] Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L., and Hong, J. Teaching johnny not to fall for phish. *Accepted in Association for Computing Machinery's Transactions on Internet Technology (TOIT)* (2009).

[112] Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., and Hong, J. Teaching johnny not to fall for phish. Tech. rep., Cranegie Mellon University, 2007. http://www.cylab.cmu.edu/files/cmucylab07003.pdf.

[113] Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., and Hong, J. Lessons from a real world evaluation of anti-phishing training. *e-Crime Researchers Summit, Anti-Phishing Working Group* (October 2008).

[114] Lee, J., Kim, J., and Moon, J. Y. What makes internet users visit cyber stores again? key design factors for customer loyalty. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2000), pp. 305–312.

[115] Lee, M. K. O., and Turban, E. A trust model for consumer internet shopping. *International Journal of Electronic Commerce 6*, 1 (2001), 65 – 91.

[116] Lininger, R., and Vines, R. D. *Phishing: Cutting the Identity Theft Line.* Indianapolis, Indiana, USA, 2005.

[117] Liu, D., Asgharpour, F., and Camp, L. J. Risk communication in computer security using mental models. *Usable Security (USEC'07)* (2007). Retrieved April 1, 2007, http://usablesecurity.org/papers/liu.pdf.

[118] Lively, C. Psychological based social engineering. Tech. rep., SANS Institute, 2003.

[119] Macmillan, N. A., and Creelman, C. D. *Detection Theory: A User's Guide.* Lawrence Erlbaum, 2004.

[120] Mail Frontier. Mailfrontier phishing IQ test, 2006. Retrieved Sept 2, 2006, http://survey.mailfrontier.com/survey/quiztest.html.

[121] Maldonado, H., Lee, J.-E. R., Brave, S., Nass, C., Nakajima, H., Yamada, R., Iwamura, K., and Morishima, Y. We learn better together: enhancing elearning with emotional characters. In *CSCL '05: Proceedings of th 2005 conference on Computer support for collaborative learning* (2005), International Society of the Learning Sciences, pp. 408–417.

[122] Mandl, H., and Levin, J. R. *Knowledge Acquisition from Text and Pictures.* North - Holland, 1989.

[123] MARKOFF, J. Larger prey are targets of phishing. News article, April 2008.

[124] Masone, C. P. Attribute-based, usefully secure email. Tech. Rep. TR2008-633, Dartmouth College, Computer Science, Hanover, NH, August 2008.

[125] Mathan, S. A., and Koedinger, K. R. *Artificial Intelligence in Education: Shaping the Future of Learning Through Intelligent Technolgis.* IOS Press, 2003, ch. Recasting the Feedback Debate: Benefits of Tutoring Error Detection and Correction Skills, pp. 13–20.

[126] Mathan, S. A., and Koedinger, K. R. Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist 40*, 4 (2005), 257–265.

[127] Mayer, R. C., Davis, J. H., and Schoorman, D. F. An integrative model of organizational trust. *The Academy of Management Review. 1995*, 3 (July 1995), 709–734.

[128] Mayer, R. E. Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology 81*, 2 (June 1989), 240–246.

[129] Mayer, R. E. *Multimedia Learning.* New York Cambridge University Press, 2001.

[130] Mayer, R. E., and Anderson, R. B. The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology 84*, 4 (December 1992), 444–452.

[131] McBride, C. M., Emmons, K. M., and Lipkus, I. M. Understanding the potential of teachable moments: the case of smoking cessation. *Health Education Research 18*, 2 (2003), 156 – 170.

[132] Mcknight, D. H., Choudhury, V., and Kacmar, C. Trust in e-commerce vendors: a two-stage model. In *Proceedings of the twenty first international conference on Information systems* (Atlanta, GA, USA, 2000), Association for Information Systems, pp. 532–536. Retrieved Sept 13, 2005, http://portal.acm.org/citation.cfm?id=359640.359807.

[133] McKnight, H. D., and Chervany, N. L. What Trust Means in E-Commerce Customer Relationships: An Interdisciplinary Conceptual Typology. *International Journal of Electronic Commerce 6* (2002), 35–59.

[134] Merrienboer, J. V., de croock, M., and Jelsma, O. The transfer paradox : Effects of contextual interference on retention andtransfer performance of a complex cognitive skill. *Perceptual and motor skills 84* (1997), 784–786.

[135] MessageLabs. Maessagelabs intelligence: 2007 annual security report. Tech. rep., MessageLabs, 2007.

[136] Microsoft. Spear phishing: Highly targeted scams. Website, September 2006. http://www.microsoft.com/athome/security/email/spear_phishing.mspx.

[137] Microsoft Corporation. Consumer awareness page on phishing, 2006. Retrieved Sep 10, 2006. http://www.microsoft.com/athome/security/email/phishing.mspx.

[138] Miller, R. C., and Wu, M. Fighting Phishing at the User Interface. *O'Reilly* (August 2005). In Lorrie Cranor and Simson Garfinkel (Eds.) Security and Usability: Designing Secure Systems that People Can Use.

[139] Mitnick, K. D., and Simon, W. L. *The Art of Deception: Controlling the Human Element of Security.* October 17, 2003.

[140] Moore, T. *Cooperative attack and defense in distributed networks.* PhD thesis, University of Cambridge, 2008.

[141] Moore, T., and Clayton, R. An empirical analysis of the current state of phishing attack and defence. In *Workshop on the Economics of Information Security* (2007).

[142] Moore, T., and Clayton, R. Examining the impact of website take-down on phishing. *e-Crime Researchers Summit, Anti-Phishing Working Group* (October 2007).

[143] Moreno, R., and Mayer, R. E. Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology 91* (1999), 358–368.

[144] Moreno, R., and Mayer, R. E. Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology 92*, 4 (December 2000), 724–733.

[145] Moreno, R., Mayer, R. E., Spires, H. A., and Lester, J. C. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction 19*, 2 (2001), 177–213.

[146] Morgan, M. G., Fischhoff, B., Bostrom, A., and Atman, C. J. *Risk Communication: A Mental Models Approach.* July 2001.

[147] Morin, R. A., and Fernandez Suarez, A. Risk aversion revisited. *Journal of Finance 38*, 4 (September 1983), 1201–16.

[148] Mutz, D. C. Social trust and e-commerce, experimental evidence for the effects of social trust on individuals' economic behavior. *Public Opinion Quarterly 69*, 3 (2005), 393–416. Retrieved Feb 20, 2006, http://poq.oxfordjournals.org/cgi/reprint/69/3/393.

[149] MySecureCyberspace. Uniform resource locator (URL), 2007. Retrieved Feb 4, 2007, http://www.mysecurecyberspace.com/encyclopedia/index/uniform-resource-locator-url-.html.

[150] Netcraft. Netcraf, 2006. Retrieved Nov 3, 2006, http://toolbar.netcraft.com/.

[151] New York State Office of Cyber Security & Critical Infrastructure Coordination. Gone phishing... a briefing on the anti-phishing exercise initiative for new york state government. Aggregate Exercise Results for public release., 2005.

[152] Nielsen, J. User education is not the answer to security problems, October 2004. http://www.useit.com/alertbox/20041025.html.

[153] NIST. Information technology security training requirements: A role- and performance-based model (800-16). Tech. rep., National Institute of Standards and Technology, 1998.

[154] NIST. Nist special publication 800-12: An introduction to computer security - the nist handbook. Tech. rep., National Institute of Standards and Technology, 2004.

[155] Patrick, A. S., Briggs, P., and Marsh, S. Designing Systems That People Will Trust. *O'Reilly* (Aug, 2005), 75–100. In Lorrie Cranor and Simson Garfinkel (Eds.) Security and Usability: Designing Secure Systems that People Can Use.

[156] Patrizio, A. Vishing Joins Phishing as Security Threat, July 2006. http://www.internetnews.com/security/article.php/3619086.

[157] Pollitt, M. G. The economics of trust, norms and networks. *Business Ethics - A European Review 11*, 2 (2002), 119–128.

[158] Ramzan, Z. Phishing attacks in and around april through september 2006. Tech. rep., Symantec, November 2006. http://www.symantec.com/avcenter/reference/phishing-stats.pdf.

[159] Reason, J. *Human Error*. Cambridge University Press, USA, October 1990.

[160] Research, G. Gartner survey shows phishing attacks escalated in 2007; more than $3 billion lost to these attacks.

[161] Resnick, P., and Zeckhauser, R. Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system. Retrieved Feb 20, 2006, http://www.si.umich.edu/ presnick/papers/ebayNBER/RZNBERBodegaBay.pdf.

[162] Riegelsberger, J., Sasse, M. A., and mccarthy, J. D. Shiny happy people building trust?: photos on e-commerce websites and consumer trust. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems* (2003).

[163] Riegelsberger, J., Sasse, M. A., and McCarthy., J. D. The Mechanics of Trust: A Framework for Research and Design. *International Journal of Human-Computer Studies 62*, 3 (2005), 381–422.

[164] Robila, S. A., and Ragucci, J. W. Don't be a phish: steps in user education. In *ITICSE '06: Proceedings of the 11th annual SIGCSE conference on Innovation and technology in computer science education* (New York, NY, USA, 2006), ACM Press, pp. 237–241. DOI=http://doi.acm.org/10.1145/1140124.1140187.

[165] Rousseau, D. M., Sitkin, S. B., Burt, R. S., and Camerer, C. Not so different after all: A cross-discipline view of trust. *The Academy of Management Review 23*, 3 (July 1998), 393 – 404.

[166] Rubin, D. C., and Wenzel, A. E. One hundred years of forgetting : A quantitative description of retention. *Psychological Review 103*, 4 (1996), 734–760.

[167] Ruyter, K. D., Wetzels, M., and Kleijnen, M. Customer adoption of e-service: an experimental study. *International Journal of Service Industry Management 12*, 2 (May 2001), 184 – 207.

[168] Salden, R., Aleven, V., Renkl, A., and Schwonke, R. Worked examples and tutored problem solving: redundant or synergistic forms of support? In *Annual Meeting of the Cognitive Science Society* (2008). In press.

[169] Salkind, N. J. *Encyclopedia of Measurement and Statistics.* Sage Publications, 2006.

[170] Salovey, P., and Rothman, A. *Social Psychology of Health.* Psychology Press, 2003.

[171] SANS. SANS top-20 2007 security risks. Tech. rep., SANS, 2007. Retrieved Dec 17, 2007, https://www2.sans.org/top20/2007/top20.pdf.

[172] Schank, R. C. Every curriculum tells a story. Tech. rep., Socraticarts.

[173] Schechter, S. E., Dhamija, R., Ozment, A., and Fischer, I. The emperor's new security indicators. In *SP '07: Proceedings of the 2007 IEEE Symposium on Security and Privacy* (Washington, DC, USA, 2007), IEEE Computer Society, pp. 51–65.

[174] Schmidt, R. A., and Bjork, R. A. New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science 3*, 4 (July 1992), 207–217.

[175] Schneier, B. Inside risks: semantic network attacks. *Commun. ACM 43*, 12 (2000), 168.

[176] Schneier, B. Semantic attacks: The third wave of network attacks. Crypto-Gram Newsletter, October 2000. Retrieved Sept 2, 2006, http://www.schneier.com/crypto-gram-0010.html#1.

[177] Schwartz, D. L., and Bransford, J. D. A time for telling. In *Cognition & Instruction* (1998), vol. 16, pp. 475–522.

[178] Scott, C. Interpersonal trust: A comparison of attitudinal and situational factors. *Human Relations 33*, 11 (1980), 805–812.

[179] Sender Policy Framework. Sender Policy Framework, 2006. Retrieved Jan 21, 2007, http://www.openspf.org/.

[180] Sheng, S. Phishing countermeasures: A public policy analysis. 2008. Ph.D. Thesis proposal.

[181] Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., and Nunge, E. Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish. In *SOUPS '07: Proceedings of the 3rd symposium on Usable privacy and security* (New York, NY, USA, March 2007), ACM, pp. 88–99. Symposium On Usable Privacy and Security.

[182] Sime, J. A., and Leitch, R. *The Nature of Expertise.* Lawrence Erlbaum Associates, 1988, ch. A learning environment based on multiple qualitative models.

[183] Singley, M., and Anderson, J. R. *The Transfer of Cognitive Skill.* Harvard University Press, USA, May 1989.

[184] Skoudis, E. Thinking fast-flux: New bait for advanced phishing tactics. *SearchSecurity* (2008).

[185] SpamAssasin. SpamAssasin, 2006. Retrieved Sept 2, 2006, http://spamassassin.apache.org/.

[186] Spence, M. Job market signaling. *Quarterly Journal of Economics 87*, 3 (1973), 355–374.

[187] SpoofGuard. Spoofguard, 2006. Retrieved Sept 2, 2006, http://crypto.stanford.edu/SpoofGuard/.

[188] SpoofStick. Spoofstick, 2006. Retrieved Sept 2, 2006, http://www.spoofstick.com/.

[189] Stanford, J., Tauber, E. R., Fogg, B., and Marable, L. Experts vs. Online Consumers: A Comparative Credibility Study of Health and Finance Web Sites., 2002. Retrieved Sept 13, 2005, http://www.consumerwebwatch.org/dynamic/web-credibility-reports-experts-vs-online.cfm.

[190] Stigler, G. J. The economics of information. *Journal of Political Economy 69*, 3 (June 1961), 213–225.

[191] Tan, Y. H., and Thoen, W. An Outline of a Trust Model for Electronic Commerce. *Applied Artificial Intelligence 14*, 8 (2000).

[192] Timko, D. The social engineering threat. *Information Systems Security Association Journal* (2008).

[193] Tversky, A., and Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *Science 185*, 4157 (1974), 1124–1131. Retrieved Sept 13, 2006, http://www.sciencemag.org/cgi/content/abstract/185/4157/1124.

[194] Tversky, A., and Shafir, E. The disjunction effect in choice under uncertainty. *American Psychological Society 3*, 5 (September 1992), 305 – 309.

[195] Ward, M. Criminals exploit net phone calls. News article, July 2006. Retrieved Dec 17, 2007, http://news.bbc.co.uk/2/hi/technology/5187518.stm.

[196] Westin, A., and Harris Louis & Associates. Health Information Privacy Survey. *Harris Louis & Associates* (1993). Conducted for Equifax Inc. 1,000 adults of the national public.

[197] Westin, A., and Interactive, H. IBM-Harris Multi- National Consumer Privacy Survey for IBM. Approximately 5,000 adults of the U.S Britain and Germany. Tech. rep., 1999.

[198] Whitten, W. B., and Bjork, R. A. Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior 16*, 4 (August 1977), 465–478.

[199] Winkler, I. S., and Dealy, B. Information security technology?...don't rely on it: a case study in social engineering. In *SSYM'95: Proceedings of the 5th conference on USENIX UNIX Security Symposium* (Berkeley, CA, USA, 1995), USENIX Association.

[200] Wogalter, M. S. *Handbook of Warnings.* Lawrence Erlbaum Associates, 2006, ch. Purposes and Scope of Warnings, pp. 3 – 9.

[201] Wogalter, M. S. *Handbook of Warnings.* Lawrence Erlbaum Associates, 2006, ch. Communication-Human Information Processing (C-HIP) Model, pp. 51 – 61.

[202] Wogalter, M. S., GodFrey, S. S., Desaulniers, G. A. F. D. R., Rothstein, P. R., and Laughery, K. R. Effectiveness of warnings. *Human Factors 29* (1987), 599 – 612.

[203] Woods, S., Hall, L., Wolke, D., Dautenhahn, K., and Sobral, D. Animated characters in bullying intervention. eCircus. Retrieved Nov 4, 2006, http://homepages.feis.herts.ac.uk/ comqkd/Woods-iva03.pdf.

[204] Wu, M. *Fighting Phishing at the User Interface.* PhD thesis, MIT, 2006. Retrieved Nov 5, 2006, http://groups.csail.mit.edu/uid/projects/phishing/minwu-thesis.pdf.

[205] Wu, M., Miller, R. C., and Garfinkel, S. L. Do Security Toolbars Actually Prevent Phishing Attacks? *Conference on Human Factors in Computing Systems (CHI)* (2006). Retrieved Feb 10, 2006, http://www.simson.net/ref/2006/CHI-security-toolbar-final.pdf.

[206] Xiao, J., Stasko, J., and Catrambone, R. Embodied conversational agents as a ui paradigm: A framework for evaluation.

[207] Yahoo. DomainKeys: Proving and Protecting Email Sender Identity, 2007. Retrieved Jan 21, 2007, http://antispam.yahoo.com/domainkeys.

[208] Ye, Z. E., and Smith, S. Trusted paths for browsers. In *Proceedings of the 11th USENIX Security Symposium* (Berkeley, CA, USA, 2002), USENIX Association, pp. 263–279.

[209] Zhang, Y., Egelman, S., Cranor, L., and Hong, J. Phinding phish: Evaluating anti-phishing tools. In *14th Annual Network and Distributed System Security Symposium* (2007). http://lorrie.cranor.org/pubs/ndss-phish-tools-final.pdf.